AI TR
Tsinghua

人工智能之人机交互

Research Report of Human-Computer Interaction

2020年第3期

Research Report of Human-Computer Interaction



清华大学人工智能研究院

北京智源人工智能研究院

清华-中国工程院知识智能联合研究中心

清华大学人工智能研究院

北京智源人工智能研究院

清华 - 中国工程院知识智能联合研究中心

2020年6月

VISIT…

LANZAROTE Caliente.COM

扫码免费领取资料



房中术(www.zuihaoziyuan.com)专注于男性增大增粗增长、阳痿、早泄、壮阳、延时、强肾、回春、健身。女性缩阴、丰胸、减肥、化妆、瑜伽、保养、产后修复、盆底肌锻炼。两性健康,夫妻按摩,房中术,性姿势,性技巧,性知识等

更多免费教程：英语学习，技能提升，PS 教学，投资赚钱，音乐教程，口才教学，情商提升，风水教学，心理学，摄影知识，幼儿学习，书法学习，记忆力提升等等......

扫码加微信lax988988领取资料



===

全站课程下载

课程不断增加

本站现资源容量已超 10T

入群联系 QQ：167520299 或添加微信：1131084518（备注PDF）

阳痿早泄训练
皇室洗髓功视频教学
女人驻颜术
泡妞约炮万元课程
足疗养 SPA 教材
玉蛋功
马氏回春功
房中术张丰川
哲龙全套视频
增大盼你增大
国际男优训练
亚当德永早泄训练
洗髓功真人内部
皇室养生绝学道家洗髓功
【铁牛人会员课】男人必备技能，理论讲解
实战高清视频
随意控制射精锻炼 视频+图片+文字
价值 1440 元第一性学名著<素女经房中养生
宝典视频>12 部
洗髓功修炼方法视频教学
陈见玉蛋功视频教学 女性缩阴锻炼
男性自然增大增长指南
强性健肾保健操 1-4
道家强肾系统锻炼功法
马氏回春功
12 堂课，全面掌握男性健康问题 让你重燃自信
联系微信：1131084518

1、东方性经
2、印度 17 式
3、口交技巧 3 部
4、港台性姿势 3 部
5、365 性姿势 6 部
6、泰国性爱密经 17 式
7、花花公子性技巧 6 部
8、阁楼艳星性技巧 7 部
9、古今鸳鸯秘谱全集 7 部
10、夫妻爱侣情趣瑜伽 2 部
11、古代宫廷性保健系列 14 部
12、汉唐宋元明清春宫图真人
13. 柔软性爱宝典 日本 9800 课
14. 李熙墨 3999 全套课
15. 妖精性爱课 2888
16. 李银河全套性课
17. 领统统性课
18. 德勇男性篇
19. 德勇男性篇
20. 缓慢性爱
21. 亚当多体位搭配篇
22. 亚当多体位结合篇
23. 德勇克服早泄讲座练习
24. 德勇以女性为中心得爱抚
25. 加藤鹰接吻爱抚舌技
26. 加藤鹰指技
27. 加藤鹰四十八手入门
28. 佐藤潮吹教学
29. 佐藤男人体能锻炼+保健品介绍
30. 佐藤男人早泄对抗训练
31. 阿拉伯延时训练

32. 田渊正浩秘籍
33. 异性性快感集中训练教学
34. 自我愉悦锻炼密宗
35. 铁牛全套延时训练课
36. Pc 机锻炼真人视频教学
37. 印度性经全集 8 部
38. 21 世纪性爱指南
39. 香蕉大叔男女训练馆全套
40. 中美真人性治疗教学+理论
41. 女性闺房秘术
42. 幸福玛利亚性课
43. 陈见如何释放性魅力征服
44. 性爱技巧讲座全套
45. 性爱秘籍全套
46. 性爱误区讲座
47. 性病讲解大全
48. 性博士讲座合集
49. 性健康和性高潮合集
50. 性教育讲座合集
51. 性能力课堂合集
52. 性生活问题解析合集
53. 意外怀孕和避孕处理课堂
54. 性感地带探索
55. 性技巧讲座
56. 性健康与性卫生讲座
57. 性生活专家答疑
58. 性心理与性道德合集
59. 性爱宝典合集
60. 性爱技巧合集
61. 完美性爱演示
62. 完美性爱技术讲解
更多精品等你来解锁哦...........

CONTENTS

1 概述篇 1
1.1 人机交互的概念..1
1.2 人机交互的发展历程..2
1.3 专家 AI TIME《论道人机交互 VS智能》 5
2 技术篇 13
2.1 触控交互..13
2.2 声控交互..15
2.3 动作交互..21
2.4 眼动交互..27
2.5 虚拟现实输入..31
2.6 多模态交互..40
2.7 信息无障碍中的智能交互技术..42
2.8 人机交互领域必读论文..46
3 人才篇 53
3.1 学者情况概览..53
3.2 代表性学者简介..56
4 应用篇 73
4.1 智能终端..73
4.2 智能穿戴..75
4.3 智能家居..76
4.4 游戏领域..77
4.5 教育领域..78
4.6 医学领域..80
5 趋势篇 83
6 结语 85
参考文献 87

AMiner

图目录

图 1-1 人机交互界面的发展 2
图 1-2 Wonder Painter示例....7
图 1-3 全手型感应 9
图 2-1 电阻式触摸屏结构示意图 13
图 2-2 电容式触摸屏原理示意图 14
图 2-3 红外触摸屏原理示意图 14
图 2-4 表面声波式触摸屏原理示意图 15
图 2-5 语音识别系统的主要模块 16
图 2-6 语音合成方法 19
图 2-7 指关节跟踪示意图 25
图 2-8 基于眼动跟踪的人机交互研究框架 28
图 2-9 不同注视位置触发示意图 30
图 2-10 (a)单行程眼势；(b)多行程眼势....31
图 2-11 用户在虚拟现实中使用 QWERTY实体键盘进行输入....32
图 2-12 (a) TipText；(b) BiTipText....36
图 2-13 VISAR键盘 38
图 2-14 小型触摸表面 39
图 2-15 (a)PizzaText；(b)RingText....40
图 3-1 人机交互领域全球学者分布 54
图 3-2 人机交互领域 TOP5国家....54
图 3-3 人机交互领域学者 h-index分布....55
图 3-4 人机交互领域中国学者分布 55
图 3-5 中国与其他国家的合作论文数量情况 56
图 5-1 人机交互技术趋势 83
图 5-2 人机交互国家趋势 84
图 5-3 人机交互机构趋势 84

表目录

表 1-1 各种人机交互界面的特征比较 4
表 3-1 近三年高产学者百人名单 70

AMiner

1概述篇

1.1人机交互的概念

人机交互（Human-Computer Interaction, HCI），作为一个术语，首次使用是在由 Stuart K. Card，Allen Newell和 Thomas P. Moran撰写的著作“The Psychology of Human-Computer Interaction”里[1]，它是一门研究系统与用户之间的交互关系的学问。系统可以是各种各样的机器，也可以是计算机化的系统和软件。人机交互界面通常是指用户可见的部分，用户通过人机交互界面与系统交流，并进行操作。人机交互技术是计算机用户界面设计中的重要内容之一，它与认知学、人机工程学、心理学等学科领域有密切的联系。

目前关于人机交互的定义主要有三种：一是 ACM（Association for Computing Machinery）的观点，它将人机交互定义为：有关交互计算机系统设计、评估、实现以及与之相关内容的学科；二是伯明翰大学教授 Alan Dix 的观点：人机交互指的是研究人、计算机以及它们之间相互作用方式的学科，学习人机交互的目的是使计算机技术更好地为人类服务；三是宾夕法尼亚州立大学 John M. Carroll 的观点：人机交互指的是有关可用性的学习和实践，是关于理解和构建用户乐于使用的软件和技术，并能在使用时发现产品有效性的学科。无论是哪一种定义方式，人机交互所关注的首要问题都是人与计算机之间的关系问题。

人机交互技术的发展与国民经济发展有着直接的联系，它是使信息技术融入社会、深入群体，达到广泛应用的技术门槛。任何一种新交互技术的诞生，都会带来其新的应用人群、新的应用领域，带来巨大的社会经济效益。从企业的角度，改善人机交互能够提高员工的生产效率，学习人机交互能够降低产品的后续支持成本。在个人的角度，可以帮助用户有效地降低错误发生的概率，避免由于错误引发的损失。在现代和未来的社会里，只要有人利用通信、计算机等信息处理技术进行社会活动，人机交互都是永恒的主题，鉴于它对科技发展的重要性，人机交互是现代信息技术、人工智能技术研究的热门方向[2]。

1.2人机交互的发展历程

人机交互的发展历史，是从人适应计算机到计算机不断地适应人的发展史，交互的信息也由精确的输入输出信息变成非精确的输入输出信息。随着网络的普及和无线通讯技术的发展，人们的需求不再局限于界面美学形式的创新，现在的用户更多的希望在使用多媒体终端时，有着更便捷、更符合他们的使用习惯，同时又有着比较美观的操作界面[3]。在过去的几十年间，人机界面经历了从命令行界面到图形用户界面两个主要发展阶段的演变；近年来，人机界面的发展越来越强调交互的自然性，即用户的交互行为与其生理和认知的习惯相吻合，随之出现的主要的交互界面形式为触摸交互界面和三维交互界面[4]。

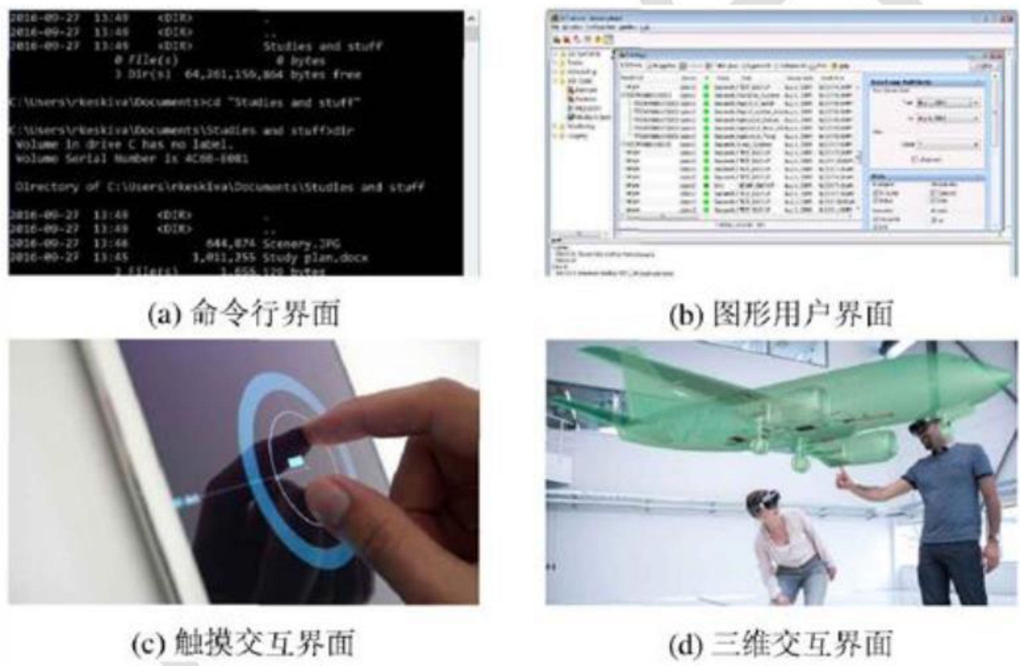


图 1-1 人机交互界面的发展

□ 命令行界面

基于命令行界面（Command-line Interface, CLI），用户使用键盘按照一定的规则输入字符，以形成可供机器识别的命令和参数，并触发计算机进行执行。其优点是由于键盘输入相对较高的准确率，以及几乎不需要冗余的操作，所以熟练的用户可以达到非常高的交互效率，同时，通过规则的设计，命令行界面也能支持丰富灵活的指令形式。但是，命令行界面的缺点在于交互非常不直观，由于机

器命令与自然语言的构造规则往往相去甚远，所以用户需要记忆大量的指令，有时甚至需要具备计算机领域的专业知识和技能，才能达到较高的使用效率。这对于新手用户而言大大提升了学习成本，也显著影响了普通用户使用命令行界面时的体验。

□ 图形用户界面

为了改进命令行界面的问题，人们提出了图形用户界面（Graphical User Interface, GUI），该界面将命令和数据以图形的方式展示给用户，用户通过所见即所得（What You See Is What You Get, WYSIWYG）的方式与显示的界面元素进行交互。根据人机交互领域中的定义，图形用户界面一般包括窗口（Window）、图标（Icon）、菜单（Menu）和指针（Pointer）这四类主要的交互元素。用户通过控制指针来对窗口、图标和菜单等显示元素进行指点（Pointing）操作，从而完成交互任务。广义的图形用户界面泛指一切用图形表征程序命令和数据的界面系统，但在狭义上，图形用户界面一般指个人电脑（PC）上的二维 WIMP界面。此时，用户与界面交互的设备一般是键盘和鼠标。图形用户界面的一大优势是摆脱了抽象的命令，通过利用人们与物理世界交互的经验来与计算机交互，从而显著降低了用户的学习和认知成本。然而，由于图形用户界面的基本操作是指点，即用户需要使用指针来选择交互目标，因而其往往对用户指点操作的精度有较高的要求。此外，由于鼠标设备所在的控制域（Motor Space）与界面显现的显示域（Visual Space）是分离的，因而用户需要对目标进行间接的交互操作（Indirect Manipulation），从而更加增加了交互的难度。

□ 触摸交互界面

在触摸交互界面上（Touch User Interface），用户通过手指在屏幕上直接操作显示的交互内容。根据人机交互研究中的定义，触摸交互界面一般包括页面（Page）、控件（Widget）、图标（Icon）和手势（Gesture）这四类主要的交互元素 A 用户通过触摸、长按、拖拽等方式直接操控手指接触的目标，或者通过绘制手势的方式触发交互指令。目前，触摸界面主要存在于智能手机和可穿戴设备（如智能手表）等设备上。触摸交互界面的优势是充分利用了人们触摸物理世

界中物体的经验，将间接的交互操作转化为直接的交互操作（Direct Manipulation），从而在保留了一部分触觉反馈的同时，进一步降低了用户的学习和认知成本。然而，触摸操作受困于著名的“胖手指问题”，即由于手指本身的柔软，以及手指点击时对于屏幕显示内容的遮挡，在触屏上点击时往往难以精确地控制落点的位置，输入信号的粒度远远低于交互元素的响应粒度。同时，由于触摸交互界面的形态仍然为二维界面，所以这限制了一些与三维交互元素的交互操作。

□ 三维交互界面

三维交互界面（3D User Interface）的出现进一步提升了人机界面的自然性。在三维交互界面中，用户一般通过身体（如手部或身体关节）做出一些动作（如空中的指点行为，或者肢体的运动轨迹等），以与三维空间中的界面元素进行交互，计算机通过捕捉用户的动作并进行意图推理，以触发对应的交互功能。目前，三维交互界面主要存在于体感交互、虚拟现实、增强现实等交互场景中。三维交互界面的优势是进一步突破了二维交互界面的限制，将交互扩展到三维空间中。因此，用户可以按照与物理世界中相同的交互方式，与虚拟的三维物体进行交互，从而进一步提升交互自然度，降低学习成本。不过，三维交互的挑战在于由于完全缺乏触觉反馈，所以用户动作行为中的噪声相对较大，而且交互动作与身体的自然运动较难区分，因而输入信号的信噪比相对较低，较难进行交互意图的准确推理，限制了交互输入的准确度。此外，由于相对于图形用户界面和触摸交互界面，动作交互的幅度一般较大，所以交互的效率也较低，同时更容易让用户感到疲劳。

表 1-1 各种人机交互界面的特征比较

交互界面	交互接口尺寸	触觉反馈	输入精度	交互效率	自然性
命令行界面	大	有	高	高	低
图形用户界面	大	有	中	中	中
触摸交互界面	小	部分	较低	较低	较高
三维交互界面	大	无	低	低	高

AMiner

每日报告

不要错过让你洞察整个商业世界的每日报告

如何免费入群？扫码加好友后回复【入群】

每日精选3份最值得学习的资料给您，不定期分享顶级外文期刊



撩她！撩她！

每日报告

不要错过让你洞察整个商业世界的每日报告

如何免费入群？扫码加好友后回复【入群】

每日精选3份最值得学习的资料给您，不定期分享顶级外文期刊



撩她！ 撩她！

每日报告

不要错过让你洞察整个商业世界的每日报告

如何免费入群？扫码加好友后回复【入群】

每日精选3份最值得学习的资料给您，不定期分享顶级外文期刊



撩她！撩她！



表 1-1汇总比较了几种交互界面的特点，可以看出，随着交互界面的演变，交互的自然性逐渐提高，但由于交互接口尺寸的限制和触觉等反馈信道的受限，导致了输入的精度和交互效率反而逐渐降低。这种交互自然性和高效性之间的制约关系，成为了人机交互研究中的难题，如何在两者之间兼顾和平衡，是具有重要理论和实践意义的研究问题。

1.3专家 AI TIME《论道人机交互 VS智能》

2019 年，清华大学人工智能研究院长张钹院士、唐杰教授、李涓子教授等人联合发起“AI TIME”science debate，希望用辩论的形式，探讨人工智能和人类未来之间的矛盾，探索人工智能领域的未来。AI TIME是一群关注人工智能发展，并有思想情怀的青年人创办的圈子。AI TIME旨在发扬科学思辨精神，邀请各界人士对人工智能理论、算法、场景、应用的本质问题进行探索，加强思想碰撞，打造成为北京乃至全国知识分享的聚集地。

AI TIME第七期《论道人机交互 VS智能》中邀请到了清华大学计算机科学与技术系长聘教授史元春、中科院软件所研究员田丰、中科院计算所研究员陈益强、小小牛创意科技 CEO曹翔等四位重量级嘉宾，论道人机交互的发展及未来，现将其探讨内容整理如下。

1.3.1人机交互的先驱学者

1979 年，当年轻的乔布斯拜访施乐 PARC 研究中心时，他看到了施乐新发明的图形用户界面（GUI），相比当时的文本命令行界面，程序图标、窗口化、下拉菜单和绚丽的图像效果把乔布斯狠狠地震撼了。“仿佛蒙在我眼睛上的纱布被揭开了一样”，《乔布斯传》中如此描述乔布斯当时的感受：“我看到了计算机产业的未来”。乔布斯回到苹果后，迅速将 GUI 移植到苹果产品上，随后上市的 Macintosh取得了轰动效果。

AMiner

今天我们习以为常的触摸屏或图形用户界面，在 40年前尚是石破惊天的革命。在计算机横空出世的这近半个多世纪里，有一些学者做出了巨大的、直接的贡献。

- **Vannevar Bush**（范内瓦·布什）

1945年，在电子计算机尚未“出世”时，范内瓦·布什就发表了题为“As We May Think”的文章，形象描述了未来个人电脑，一种被称为 MEMEX的机器，阐释了直接交互、超链接、网络存储等概念。

- **J.C.R Licklider**（约瑟夫·利克莱德）

1960 年，约瑟夫·利克莱德提出“人机共生”的思想，并在布什的领导下通过美国国家科技计划大力支持了人机共生理念下的图形与可视化、虚拟对象操控、互联网络等研究项目，在他的主导下，个人电脑、互联网络的标志性关键技术在六七十年代逐次诞生了。约瑟夫·利克莱德领导的交互式计算，不但研发了分式操作系统，而且直接地引导了图形技术。

- **Douglas C. Engelbart**（道格拉斯·恩格尔巴特）

恩格尔巴特在 1963 年设计出一款手掌大小、以轮子为基础的设备，此设备也就是鼠标的原型，1968 年在旧金山的秋季联合计算机会议上，恩格尔巴特先后演示了鼠标、所见即所得的文字编辑器、超链接、文本图形混排等，还谈到了阿帕网（ARPANet，互联网的前身）以及科技进步的未来。他是人机交互的先锋，开发了超文本系统、网络计算机，以及图形用户界面的先驱；并致力于倡导运用计算机和网络，来协同解决世界上日益增长的紧急而又复杂的问题。他被冠为“鼠标之父”。

在 Vannevar Bush、J.C.R Licklider、Douglas C. Engelbart等先驱的推动下，在语言学、心理学、计算机科学的共同参与下，计算机从没有用户界面，到有了图形用户界面，开创了个人电脑以及互联网络等惠及整个社会的新产业。未来，在新的传感和多媒体技术的共同支持下，机器将可以通过感知和数据处理技术来理解我们，来理解周围的环境，实现更自然、更智能的人机交互。

AMiner



1.3.2嘉宾分享案例

每位嘉宾就所做所见，讲述了人机交互正在或即将发生的场景，这些场景基本代表了人机交互最新、最前沿的研究方向。

- **Wonder Painter**

Wonder Painter，由北京小小牛创意科技研发，是一款融合了人工智能、机器学习和人机交互为一体的新型技术，可以将任何静态物体立即转换成动画。



图 1-2 Wonder Painter示例

曹翔老师现场展示了 Wonder Painter工作的典型例子。一张普通的纸，一支普通的笔，画下天马行空的图画，再由手机采集起来，瞬间就可以转化成三维的动画。曹翔老师希望通过这项技术，打破普通人表达创意的门槛。

- 输入技术和相关理论

自然人机交互具有输入非精确性等新型特征，同时 EMG等新的输入通道为自然人机交互提供了研究机遇。如何建立针对新型特征的运动模型，并利用生理等新型通道感知用户交互意图是人机交互研究的重点之一。田丰老师就此介绍了自己的两项工作。

一是运动目标获取非确定性模型。在运动的画面中，如何知道用户真正想点击的是哪一个模块，其基本思路是：1）构建落点分布模型，揭示落点分布与目

标初始位置、大小间的关系；2）对目标获取错误率进行预测；3）增强移动目标获取成功率。

二是基于 EMG的交互意图隐式感知技术。想象一下，你拿起一支笔，或者喝了一口水，不等你给机器传达命令，它就通过你的肌肉电信号，感知到了你现在正在做什么。这就是田丰目前在做的另一项研究。该项工作的主要思路是：1）提出一种通过肌电感知技术对日常手部动作及所接触物品属性进行感知的方法；2）通过布置手臂上的传感器捕获手部活动时的 EMG信号；3）提取能表征 EMG信号的均方根、AR系数等特征并通过机器学习方法识别物体。

□ 多模态协同感知

陈益强老师指出，普适计算时代，未来的人机交互模式应该是多模态的。我们既可以用键盘、鼠标、语音进行机器操作，也可以用手势、表情、唇语进行操作。围绕“多模态”设想，陈益强提到了自己的两项工作。

一是基于多模态感知理论的手语识别。陈益强将面部识别、手势动作识别和手语识别相融合，来提高手语识别的精度，以期帮助残障人士和外界沟通。

二是基于多模态感知理论的人机交互方法，利用语音，视觉，可穿戴等等，使机器人获知用户当前复杂行为及情感状态。陈益强提到：“人机交互的终极目标，是达到人机之间的无缝互动，仿佛人和人在交互一样。因此，机器要准确感知到人的当前动作、行为甚至情绪。我们基于多模态手段，如语音、面部表情、可穿戴生理指标检测等，试图解决这个问题。”

□ 自然用户意图的准确理解

在使用手机软键盘时，经常发生误触的问题，26 个字母挤在狭窄的输入界面里，再配上胖乎乎的手指，点错的经历太多了。这是触屏这种自然交互界面上典型的难题：胖手指难题。输入信号脱离了实体按键，通过触屏控制输入仍然可以达到精准的输入效果。

AMiner

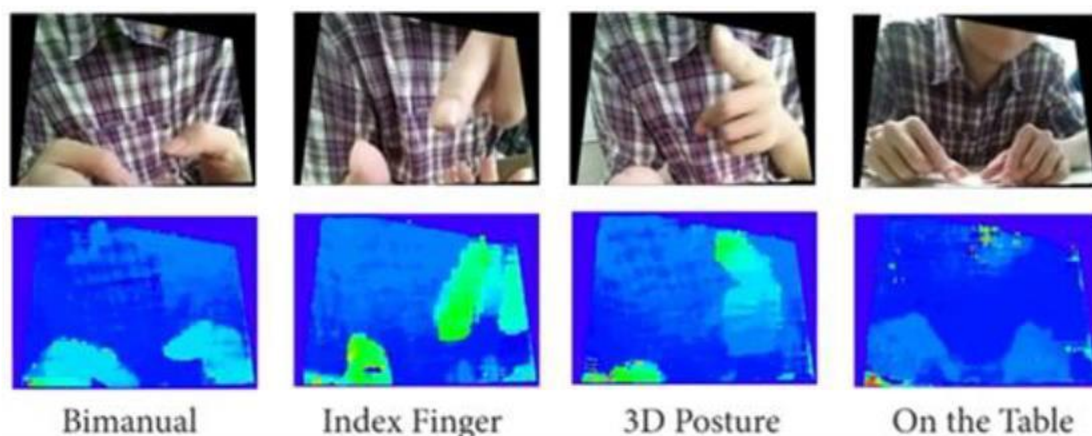


图 1-3 全手型感应

史元春教授的研究工作，提出了基于贝叶斯推理的自然用户意图理解框架，建模用户行为特征，在模糊的输入信号上推测用户的真实意图。基于这项技术，史元春教授的团队已经研究实现了手机、平板、头盔、电视等一系列接口上的输入法，输入准确度大幅度提高，且几乎不需要视觉瞄准，进而还能支持盲人用户准确实现软键盘输入。史教授正在研制的手机前置摄像头的操控行为。通过感知到人手在界面上的操作变化，可以做出新的“输入法”。比如手握手机的任意边框或位置，就可以输入信息、访问界面。这些操作无需经过人眼确认，通过字体感知即可实现准确输入。这些成果已见诸实际应用，如：电容屏防误触技术部署在了华为 MATE 系列千万量级手机上，智能输入意图推理算法应用在搜狗和华为的输入法上，无障碍手机交互技术应用于手机淘宝和支付宝等。史元春教授指出，要建立一套理解人的意图表达的计算框架，还要继续在理论和传感技术上做更多研究。

1.3.3技术思辨

嘉宾从人机交互的理论和方法、与人工智能的关系和未来技术发展趋势和挑战、人才培养等方面发表了自己的看法并进行了热烈的讨论。

□ 人机交互的理论和方法

Q1：交互界面的构建有计算模型吗？

AMiner

Q2：如何定量评估界面设计的效果？

田丰认为交互界面的构建是有计算模型的，但针对自然人机交互，研究者做了“点”上的研究，还没有完全系统化。

史元春教授同意以上观点，并指出定量评估的方法虽有，但很不充分。不过借助相应的传感技术，定量评估的原理和技术都在不断拓展，这从红外反射监测血流、血压参数、情绪变化等一系列应用上就可以看出。

曹翔补充道：人机交互很难用计算机模型衡量，是因为人机交互的任务是多样的，且越来越娱乐化，这令交互效率很难界定。我们越来越需要通过生理指标衡量人机交互的体验，这里面一定需要定量数据。明确的任务性的工作相对容易建模，因为目的很清晰；体验性的、娱乐性的、沟通性的工作比较难用计算的方式建模，因为其中夹杂着大量非简单人机交互的内容，例如人与人之间的互动等。

□ 人机交互与人工智能的关系

Q1：在中文上，两者都有关于“人”的研究，有共同的研究内容和方法吗？

Q2：两者研究成果的价值体现有何异同？

Q3：智能人机交互主要指什么？

Q4：人机交互研究对 *AI*有贡献吗？

陈益强提出：要做好智能人机交互，必须做到个性化。人脑智能分为三个部分，中枢神经、小脑和大脑，这三部分体现了不同程度的智能，可对应人机交互中的不同智能应用程度。比如神经智能，中枢神经控制下的膝跳反射或条件反射就可和键盘鼠标触屏技术等传统人机交互技术相对应，注重实时的感知与执行，实现敲一下键盘，屏幕就弹出一个字。而像语音识别、手势识别这一部分新型人机交互，就类似于小脑智能部分，侧重于基于学习或预测的执行，过程包括了感知-学习-执行。而智能人机交互的终极目标，需要在感知刺激的基础上不仅有学习，还应有知识推理，而后决策执行，这类似于大脑智能层次，到了这个部分，人机交互应该能做到带情感甚至带有价值判断的智能人机交互。

史元春教授认为，人机交互应该让机器更好的适应人，适应人的本性，适应人的操控能力、感知能力和认知能力。从“人”的研究内容上来说，人机交互与人工智能有差异，但出发点是一致的，即“人机共生”。目前看，人工智能的研究更多的体现在人的识别、语言的表达等数据密集型任务上的处理方法，人机交互的研究更偏重于对人的主动交互行为和感知能力的建模、传感和建立适应的接口技术，人机关系必定向着共生的方向发展，这些研究内容和方法会相互影响和适应，交叠的研究内容会越来越多。“做人工智能最后要接触人机交互，做人机交互最后也要接触人工智能。”

田丰提出了“人机交互和人工智能从交替沉浮到协同共进”的观点。未来的计算机将是一个智能体，人与智能计算机的交互，即是人机协同；而从人工智能的角度讲，自动驾驶、自动诊断也讲求人机协同，两者殊途同归。人机交互想解决不确定性，就需学习人工智能的方法，两者互相支撑，协同并进。

曹翔补充：机器在不断地取代人力，未来所有带“老”字的职业，都将被 AI取代，例如老司机、老中医等。为什么呢？因为 AI最擅长数据和经验，完全依靠经验驱动的职业很可能被 AI取代。创造性的、沟通性的、娱乐性的工作，则是不可被 AI取代的，这就体现出了人机交互的重要性。

陈益强不完全同意曹翔的观点，他指出：“我认为带‘老’字的职业都不会被取代。例如一台自然语言处理及知识推理能力极强的类似医疗诊断机器，它可以借助公开发表的文献（大数据）习得 80%的经验，但剩余 20%的疑难杂症（小数据）无法习得，这部分只能求助于‘老’医生。”

 如何评论未来交互技术的发展

Q1：最近出版的《科技之巅》总结了近十年全球百项突破性技术，包括 *IT*、能源、生物医药、材料等共十章。其中，人工智能与人机交互分列第一和第二章，*HCI*主要集中在手势、语音交互和穿戴产品，可否对现在人机交互技术和未来交互技术的发展做评论？

Q2：您最关心的 *HCI*挑战是什么？

史元春教授讲道："未来计算机的形态会变化，甚至可能不存在了，但计算机技术会持续为我们服务，成为人机共生的一部分，交互接口、交互任务会有很大的变化，但会更自然，更智能。"

陈益强指出，普适计算可以使手环、穿戴设备等计算机形式化于无形，就像看不见摸不着的空气一样。例如穿戴设备可以附着在衣服、鞋子里，实现人机共生。最后在材料、计算技术的进步下，真正实现对人类自然行为的意图理解，助力解决人口老龄化、阿尔茨海默病早期预警等。

曹翔最关心的是，如何通过机器、技术放大个人的创造力，从而帮助个人在社会和工作中立足。"我们五感的潜力，不只局限于物理世界的刺激。随着技术的进一步发展，我们可以挖掘更多的感官体验力，创造全新的体验。"

田丰老师更关注如何通过人机交互的研究推动相关产业的发展，产生主流的影响。例如电子白板对中国教育信息化的推动，人工智能对帕金森、脑卒中、阿尔茨海默、痴呆等做辅助诊断等。值得一提的是，田丰带领团队研发的笔式电子教学系统获得了国家科技进步二等奖，并与协和医院共同取得了国家卫健委颁发的医疗健康人工智能应用落地 30最佳案例的荣誉。

□ 人机交互的人才发展路径

Q1：工业界需要什么样的交互人才？

Q2：学校如何培养人机交互人才？

Q3：如何判断自己适不适合做人机交互？

史元春教授提到："我们培养的人才应该能够发现交互难题，并且能通过科学的方法来解决这个问题"。曹翔赞同史元春教授的观点，并补充道："交互设计师、用户研究员等对口培养的专业，不难找工作；难找工作的是把人机交互作为一个研究领域去学习的学生，因为现有的一个萝卜一个坑的职业体系，并不太适合跨学科的人才，但创业特别需要这样的"。如何判断自己适不适合做人机交互？曹翔提醒道，跟"风口"很不靠谱，兴趣才是首要的。

2技术篇

本篇基于用户与系统之间进行交流操作所主要使用的触摸、手势、语音和视觉等自然感官的顺序来进行相关技术介绍。具体包括触控交互、声控交互、动作交互、眼动交互、虚拟现实输入、多模式交互以及智能交互等人机交互技术。

2.1触控交互

显示器从仅向用户输出可视信息到成为一种交互界面装置主要是归因于触控功能与显示器的一体化模式，尤其是在移动装置上的使用。从 1965 年第一份电容触摸屏报告诞生至今[5]，经过近几十年的发展，触控式交互技术已经成功应用于全球主流消费品，触控式交互技术能让人们通过触摸就能直接与屏幕内容互动，让人们不用或进行很少的训练就能有更为便捷的使用体验。有很多不同的技术能够实现触控式交互，下面我们简要介绍电容、电阻、光学和声学交互技术[6]。

□ 电阻式触控技术

电阻触摸屏通过压力感应原理来实现对屏幕进行操作和控制。当手指触摸屏幕时，薄膜下层的 ITO会和玻璃上层的 ITO有一个接触点，在 X轴方向就其中一面导电层导通了 5V均匀电压场，此时采样得到的电压由零变为一个正电压值，感应器检测到电压导通，传出相应的电信号，进行模 /数转换，最终将转换后的电压值与 5V 相比，即可计算出触摸点的 X 轴坐标值。同理可以计算出 Y 轴的坐标值，这样就完成了点选的动作，并呈现在屏幕上。

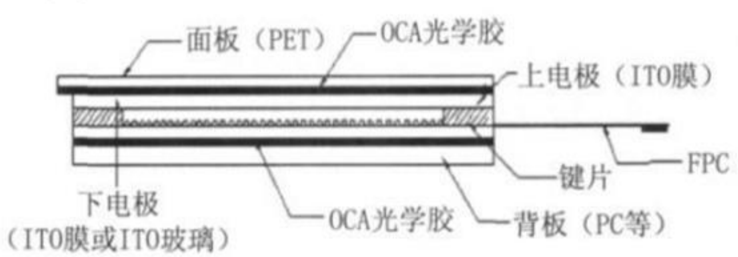


图 2-1 电阻式触摸屏结构示意图

□ 电容式触控技术

当手指触摸电容式触摸屏时，在工作面接通高频信号，此时手指与触摸屏工作面形成一个耦合电容，这相当于导体，因为工作面上有高频信号，手指触摸时在触摸点吸走一个小电流，这个小电流分别从触摸屏的四个角上的电极流出，流经四个电极的电流与手指到四角的直线距离成比例，控制器通过对四个电流比例的计算，即可得出接触点坐标值，如图 2-2所示。



图 2-2 电容式触摸屏原理示意图

□ 红外触控技术

红外触摸屏的基本结构如图 2-3所示，在外框的上下、左右两个对边上分别对应装有红外发射管和红外接收管，当手指触摸屏幕时，红外光线将被阻断，依次选通红外发射管及其对应的红外接收管，在屏幕上方形成一个红外线矩阵平面，从而致使红外接收端的电压产生变化，红外接收端的电压经过 A/D 转换送达控制端，控制端将据此进行计算得出触摸位置。

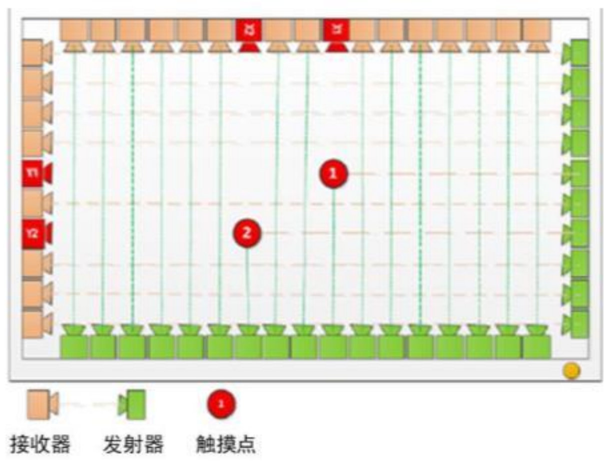


图 2-3 红外触摸屏原理示意图

 表面声波触控技术

表面声波式触摸屏主要依靠安装在强化玻璃边角上的超声波换能器来实现触摸控制的。当手指触摸显示屏时，手指阻挡了一部分声波能量的传播，此时接收波形将会发生变化，在波形图上可以看见即某一时刻波形发生衰减，通过这个衰减信号控制器就可以计算出触摸点位置。

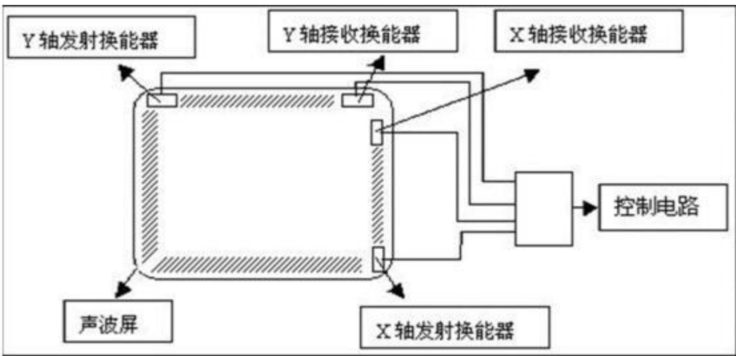


图 2-4 表面声波式触摸屏原理示意图

2.2声控交互

人与人之间最有效也是最普遍的交互形式是有声语言，虽然一般情况下我们可以毫不费力的表达和理解他人的话语，但是让一台计算机具备人类拥有的对有声语言的理解能力绝非易事，学术界和企业界对语音控制的人机交互有着浓厚兴趣并在不断探索[7]。目前，从交互显示的观点来说，声音界面似乎可以使我们更简单的与各种形态的显示器交互，从而获得更为本真的体验。下面我们简要介绍语音识别、自然语言处理及语音合成三项关键技术。

2.2.1语音识别

语音识别是将音频数据转化为文本或其他计算机可以处理的信息的技术。图2-5展示的是一个典型语音识别系统的结构，主要由 4个部分组成：特征提取、声学模型、语言模型和解码器搜索。

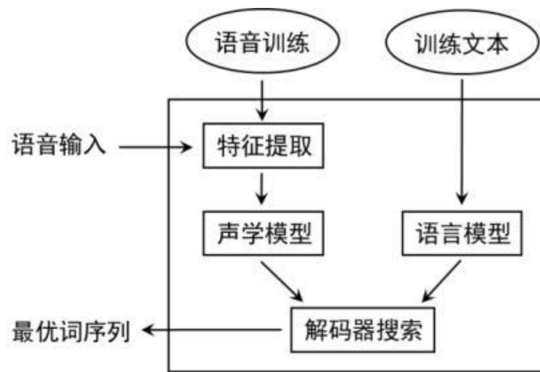


图 2-5 语音识别系统的主要模块

□ 特征提取

模拟的语音信号进行采样得到波形数据之后，首先要送入到特征提取模块，提取出合适的声学特征参数供后续声学模型训练使用。好的声学特征应当考虑以下三个方面的因素。第一，应当具有比较优秀的区分特性，以使声学模型不同的建模单元可以方便准确的建模。其次，特征提取也可以认为是语音信息的压缩编码过程，既需要将信道、说话人的因素消除保留与内容相关的信息，又需要在不损失过多有用信息的情况下使用尽量低的参数维度，便于高效准确的进行模型的训练。最后，需要考虑鲁棒性，即对环境，噪声的抗干扰能力。

当前，最常用的经典声学特征提取方法是基于倒谱分析的特征参数提取，如感知线性预测（ Perceptual Linear Prediction, PLP）系数和梅尔域倒谱系数（Mel-Frequency Cepstral Coefficients, MFCC）。除了此以外，为了提高特征参数的区分性，研究者还提出其他方法来实现特征变换和特征降维的目的，包括主分量分析、线性判别分析、异方差线性判别分析等。也出现了利用区分性准则对输入原始特征进行变换的技术[8]。

□ 声学模型

声学模型在语音识别系统中起着至关重要的作用，描述了声学基元产生特征序列的变换过程。给定一个声学特征矢量，依据声学模型来计算它属于每个基元的概率值，通过最大似然准则得出与特征序列对应的状态序列。

（1）声学基元选择

AMiner

声学基元的选择是声学建模中很关键的问题，合适粒度的基元对系统性能的提升有很大帮助。语音识别把词（Word）、音节（Syllable）、声韵母（Initial/Final）以及音素（Phone）等作为常用基元。基元的选择和设计通常考虑的是语音学知识，基元也可基于数据驱动的方法产生，虽然有些在语音学上意义不明确，但是有的效果也不错。在语音识别中为了应对协同发音现象从而能更准确的描述语音，通常选用上下文相关的方法来对语音建模，这其中有两种情况，当只考虑前一音子对当前音子的影响被称为 Bi-Phone，同时考虑前一音子和后一音子的影响被称为 Tri-Phone。对英语识别中通常采用上下文相关的三音素作为基元，对于汉语建模往往采用音节或者上下文相关的声韵母作为基元。

（2）隐马尔科夫模型

如今主流语音识别系统都采用隐马尔科夫模型（HMM）作为声学模型，这是因为 HMM具有很多优良特性。HMM模型的状态跳转模型很适合人类语音的短时平稳特性，可以对不断产生的观测值（语音信号）进行方便的统计建模；与 HMM 相伴生的动态规划算法可以有效地实现对可变长度的时间序列进行分段和分类的功能；HMM的应用范围广泛，只要选择不同的生成概率密度，离散分布或者连续分布，都可以使用 HMM进行建模。HMM以及与之相关的技术在语音识别系统中处于最核心的地位。

□ 语言模型

模型主要是刻画人类语言表达的方式习惯，着重描述了词与词在排列结构上的内在联系。在语音识别解码的过程中，在词内转移参考发声词典、词间转移参考语言模型，好的语言模型不仅能够提高解码效率，还能在一定程度上提高识别率[9]。基于 N-gram 的统计语言模型是当前统计语音识别框架下最常使用的语言模型，用来表示长度为 N 的词串的出现概率，其核心思想是用一个在各词之间进行跳转的 N-1阶马尔科夫过程来描述词串的生成过程，所以，词串的概率 $p(W)$ 可以表示为：

$$p(W) = p(w_1^K) = \prod_{k=1}^{K} p(w_k \mid w_1^{k-1}) \approx \prod_{k=1}^{K} p(w_k \mid w_{\max[k-N+1,1]}^{k-1})$$

AMiner

这里 K 表示该词序列中包含的词个数，w_k表示词序列中第 k 个词。前词的概率仅仅取决于其前面 N-1个词的历史，这也是 N-gram语言模型命名的缘由[8]。

□ 解码器

在声学特征提取完成，声学模型和语言模型训练完成以后，语音识别就是指结合声学模型和语言模型，利用相关搜索算法在解码器中找出最优词序列的过程。不难看出，如果不作任何限制，搜索空间相对于词表大小和语音中可能出现的词数目是以指数级增长的。巨大的搜索空间带来的运算量是无法想象的，也是很多实时语音识别任务要尽量避免的。解码器搜索效率的高低直接关系到语音识别系统的实用程度。所以，必须通过一些有效的优化算法，提高解码效率，把原来超大规模的搜索问题压缩到计算机可以有效处理的程度。

维特比算法（Viterbi Algorithm）是当前绝大部分主流解码器中使用一种有效的压缩搜索空间的近似方法。维特比算法是时间同步的，需要在解码过程中进行同步的快速概率计算，以及裁剪搜索空间的处理：包括快速计算输出概率的高斯选择算法、Beam裁剪算法、语言模型等。除此之外，也有一些语音识别系统的解码器使用异步的堆栈解码算法，利用一些具有启发性的度量来指引搜索算法的完成。时间异步的搜索方法，一般会使用简单的模型来迅速地生成识别结果的备选空间，然后再使用更加精细的模型对备选空间重新计算得分并生成最终的最优识别结果。识别结果的备选空间一般采用 N-best 列表，或者使用包含更多信息的词图来表示。

2.2.2语音合成

语音合成就是将一系列的输入文字信号序列经过适当的韵律处理后，送入合成器，产生出具有尽可能丰富表现力和高自然度的语音输出，从而使计算机或相关的系统能够发出像“人”一样自然流利声音的技术[10]。

语音合成的发展经历了机械式语音合成、电子式语音合成和基于计算机的语音合成发展阶段。基于计算机的合成方法由于侧重点不同，语音合成方法的分类也有差异。但主流的、获得多数认同的分类则是将语音合成方法按照设计的主要

思想分为规则驱动方法和数据驱动方法。前者的主要思想是根据人类发音物理过程从而制定一系列规则来模拟这一过程，后者则是在语音库中的数据上利用统计方法如建模来实现合成的方法，因而数据驱动方法更多的依赖语音语料库的质量、规模和最小单元等。语音合成的具体分类如图 2-6所示，各个方法也不是完全独立的，近些年来研究人员取长补短地将它们整合到一起[11]。

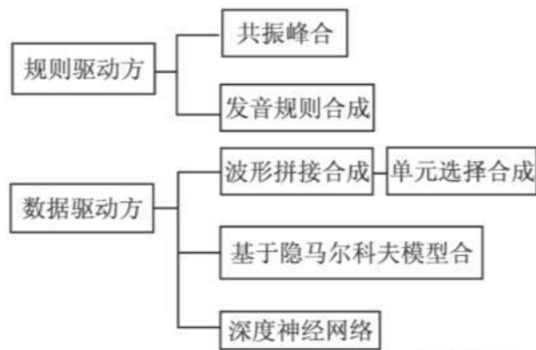


图 2-6 语音合成方法

□ 共振峰合成

共振峰是指声道的共振频率，共振峰合成是指用共振峰来加权叠加生成语音。从滤波器的观点来看，语音的产生是一个声源的激励加时变滤波的过程。脉冲发生器模拟产生浊音的声带振动激励，清音是由声带中气息的湍流噪声造成的，用一个噪声发生器来模拟。所有的语音都是这两类声源通过频率响应不同的滤波器处理后得到，用一个多通道的时变滤波器来模拟，使得其输出具有目标语音的频谱特性。经过放大器（口唇辐射）输出，就可以听到合成语音。最初，共振峰合成出的语音自然度很低，有些学者提出是因为共振峰建模时忽略了谱的变化。经过在共振峰合成中加入或改进谱建模。共振峰合成的语音的自然度被提升了，所以也常用来生成特色的语音。

□ 发音过程合成

发音过程模拟合成是直接模拟人的发音这一物理过程，通常制定一系列规则来操控模型发声。由于得到真实发音的物理过程难度大，这一方法也较难实现。但它的优点在于，一旦一个精细较为准确的规则建立，就使得这个系统有很大的可塑性和灵活性。规则驱动方法的另一不足在于对超音段的控制不足，自然度受

损，以至于出现人们难以接受的机器声音。为了在高复杂度和高自然度之间做一个平衡，研究人员采用预先录制的语音库，通过拼凑语音库单元来快速生成较高质量的语音。

- 波形拼接

波形拼接方法通过连接小的、事先录好的语音单元，如音素，双音素，三音素等并经过韵律修饰来拼接整合成完整的语音。波形拼接技术是一种通过波形处理，使得言语的超音段特征发生改变，而音段特征（谱包络）保持不变的时间维处理技术。这种技术最大限度的保留了原始发音人的音质，自然度和清晰度都很高，达到人们能够接受的水平。但这样直接拼接的方法导致语音听起来人工、生硬，韵律修饰导致边界处明显不连续。拼接处容易产生意想不到的错误，合成效果不稳定，音库容量大，构建周期长，可扩展性太差，不适宜作为嵌入式应用。但如果要合成的语句中的大部分单元都在语音库里存在，那么合成出的语音的自然度要比规则拼接高得多，以至于当寻求高自然度时，这类方法成为主流方法。但它的代价则是设计精细、科学，占用内存大，人力物力耗费巨大的语音语料库。

- 单元选择

单元选择是一种波形拼接方法，但是它在事先录好的库中存储了每个拼接单元的大量不同韵律实例，这样就避免了传统波形拼接中的韵律修饰，也就解决了传统波形拼接方法中语音单元边界不连续的问题。一般来说，单元选择方法合成的语音音质好，稳定，自然度较高。但单元选择方法也像其他波形拼接方法一样存在拼接时选择了错误单元的情况。

- 谐波加噪声模型

为了解决单元选择中的误拼情况，研究人员又提出了谐波加噪声模型，该模型将语音信号看成是各种分量谐波和噪声的加权和，对信号的这种分解使得合成出的信号更加自然。

- HMM模型和 STRAIGHT合成技术

AMiner

如前所述，波形拼接方法需要的语音语料库非常占用资源而且要求设计精细，因为它所有的拼接单元全都来自于库，而且训练模型的时间通常很长。隐马尔科夫模型（HMM）结合谐波加噪声模型一起，解决了这个问题。这种方法也被看作是最有用的统计建模方法。它的流程如下：首先，选择合适的特征表征语音库中的语音，训练模型；然后，利用模型将文本生成序列状态的特征向量；最后送入一个滤波器，将特征向量转换成语音。基于 HMM模型建模方法，灵活度高，库小，并且构建时间也少，非常适合移动嵌入式平台。20 世纪出现了数据驱动向规则驱动的倾向，其重要标注就是新的语音处理技术和 HMM统计模型，使得参数合成出现了新局面。

 神经网络及深度神经网络模型

深度神经网络属于多层神经网络，二者在结构上大致相似，不同的是深度学习网络在做有监督学习的时候先做非监督学习，然后将非监督学习到的权值当作有监督学习的初值进行训练。

2.3动作交互

2.3.1发展现状

 基于动作的目标获取技术

目标获取是人机交互过程中的最基本的交互任务，用户向计算机指明想要交互的目标，其他的交互命令均在此基础上完成。随着交互界面的发展，在很多自然交互界面上，如远距离大屏幕，虚拟现实和增强现实设备等，传统的交互设备（如鼠标，键盘）无法继续用来完成目标获取任务。因此，在这些界面上，研究者探索使用动作交互完成目标获取任务的可能方式。主要的输入方式分为直接和间接两种。直接的动作选取要求用户通过接触目标位置的方式对其进行选取，例如在增强现实应用中，用户通过以手部接触的方式完成虚拟物体的选取。间接的目标选取方式则需要用户通过身体部分的位置和姿态来控制和移动光标，再借助光标指示目标的位置进行选取。其中，一个广泛应用的光标控制方法是光线投射

方法（Ray-casting），用户通过控制一束虚拟光线来选取与之相交的目标。多种控制方式已被广泛研究，包括通过手指延伸方向，头部朝向方向和手眼连线等方式控制光线的起始位置和指向方向，进而控制光线指向想要选取的目标。在通过直接或者间接的指点方式指明要选取的目标后，目标选取技术还需要用户完成一个选取确认的过程。该过程用于避免用户无交互意识的动作被误识别为目标选取动作而引起误触发问题。因此，基于动作的目标选取方法一般需要用户做一个确认动作来完成选取过程。例如，想要在目前商用的增强现实头显设备（如微软Hololens）完成一次目标选取，在用户移动光标指向目标后，还要完成一个空中手势作为确认（如 Hololens的 Air Tap手势）。除确认动作外，相关技术也尝试使用光标暂留和基于光标轨迹的确认方式。光标暂留方式要求用户将光标移动到目标位置后维持在目标内部一段时间直到超过选取确认的时间阈值。基于光标轨迹的确认方法需要特殊设计目标的外形，同时要求额外的模式切换功能，用户切换到选取模式后控制光标穿过目标的边界将完成目标获取的确认。

□ 自然的动作命令映射方法

在向计算机指明想要交互的目标对象基础上，用户需要进一步传达想要对交互目标完成的交互意图。动作输入技术可以支持这一交互意图传达过程，方法为将一系列交互动作映射到对应的交互指令上，当用户完成其中之一的交互动作时，计算机利用预设的映射关系解码交互动作，执行对应的交互指令。而如何实现自然的动作到指令的映射关系则决定着输入技术和交互动作的可用性，可发现性等影响用户体验的因素。在以往的动作输入技术中，动作命令的映射关系由开发者或者设计师决定。这种基于经验的定义方法往往存在自然性和识别准确率难以权衡的矛盾。设计师会更加注重自然性，系统开发者更注重保证交互动作的识别效果。因此，这些动作输入技术面临着动作交互可发现性低和学习成本高的问题，这也是动作交互未能更加广泛被应用的重要原因。为了解决映射关系的自然性问题，研究者提出用户参与式的动作输入设计方法，让使用输入技术的用户本身参与到映射关系的确定过程中。该方法最早被应用在为可交互桌面设计交互动作完成界面控制的研究中。首先向用户展示交互动作将会引发的交互效果，随后要求用户去定义该交互动作的具体形式，最后统计不同用户的定义结果，选取最

高频率的交互动作对应到指定交互效果上。这样设计出的交互动作往往与用户的日常经验相关，因而有更高的可记忆性也被用户所偏好。该研究方法被成功应用到移动设备交互，智能电视交互，虚拟现实和增强现实等应用领域中。

□ 自然交互动作的识别方法

在计算机将交互动作解码为用户的交互意图之前，首先要对用户完成的交互动作进行感知和识别。计算机需要借助传感器将用户的交互动作转换为可以计算和分析的信号数据，随后对于信号数据进行分割，特征提取和分类。常用的传感信号包括图像，声音，惯性传感器信号等。基于图像的用户身体姿态感知已被广泛应用于远距离大屏幕交互中。通过使用深度摄像头（如微软 Kinect 摄像头）作为传感设备，算法可以提取出用户当前的骨架信息（Skeleton），通过感知一段时间窗口内的骨架信息变化来识别用户的交互动作。基于声音信号的事件检测也已被深入研究。用户日常活动（如开门）和紧急事件的检测（如鸣枪，尖叫等）均可通过单个或者多个麦克风采集到的音频信号来识别和分类。在将交互动作感知为连续的传感器信号后，计算机还要对信号进行分析和特征提取，以及最终的分类。常用的分类算法包括基于逻辑启发，基于数据模板和基于机器学习的分类方法。基于逻辑启发的分类方法通过设定逻辑规则来识别不同交互动作，典型工作设定动作幅度和方向的阈值，当用户的交互动作超出这些阈值时会被识别为对应的动作类别。基于数据模板的分类方法通过采集不同类别交互动作的信号数据，通过比较用户当前的动作信号与模板数据的相似程度来判断该动作的类型。基于机器学习的方法，通过提取特征来表征交互动作，并构建不同模型分类器（如支持向量机）来对交互动作进行分类[12]。

2.3.2手势识别

手势可定义为人手或者手和手臂相结合所产生的各种姿态和动作，它分为静态手势（指姿态，单个手形）和动态手势（指动作，由一系列姿态组成），前者对应模型空间里的一个点，后者对应一条轨迹。相应地，可以将手势识别分为静态手势识别和动态手势识别[13]。手势无论是静态或动态，其识别顺序首先需进

行图像的获取、手的检测和分割、手势的分析，然后进行静态或动态的手势识别。下面简单介绍手势分割、手势分析及手势识别技术[14]。

□ 手势分割

手势分割是手势识别过程中关键的一步，手势分割的效果直接影响到下一步手势分析及最终的手势识别。目前最常用的手势分割法主要包括基于单目视觉的手势分割和基于立体视觉的手势分割。

单目视觉是利用一个图像采集设备获得手势，得到手势的平面模型。常用的方法主要有：（1）基于徒手的表观特征识别（徒手的表观特征指手的肤色、纹理、指尖、手的轮廓、手型和手的大小等），通常都利用肤色信息在 YUV、HSV 或 YCbCr等颜色空间下建模对手势进行分割。（2）人为增加限制的方法，如使用黑色和白色的墙壁、深色的服装等简化背景或要求人手佩戴特殊的手套等强调前景，以简化手区域与背景区域的划分。

而立体视觉是利用多个图像采集设备得到手势的不同图像，转换成立体模型。主要方法有立体匹配和三维重构。基于立体视觉的手势分割需建立手势的三维模型，相比二维模型其所需设备较多，需要两个或多个图像采集设备获得手势图像。立体匹配的方法与单目视觉中的模板匹配方法类似，也要建立大量的手势库；而三维重构则需建立手势的三维模型，计算量将增加，但分割效果较好。

□ 手势分析

手势分析是完成手势识别系统的关键技术之一。通过手势分析，可获得手势的形状特征或运动轨迹。手势的形状和运动轨迹是动态手势识别中的重要特征，与手势所表达意义有直接的关系。手势分析的主要方法有以下几类：边缘轮廓提取法、质心手指等多特征结合法以及指关节式跟踪法等。边缘轮廓提取法是手势分析常用的方法之一，手型因其特有的外形而与其他物体区分；多特征结合法则是根据手的物理特性分析手势的姿势或轨迹；关节式跟踪法主要是构建手的二维或三维模型，再根据人手关节点的位置变化来进行跟踪，其主要应用于动态轨迹跟踪。

AMiner

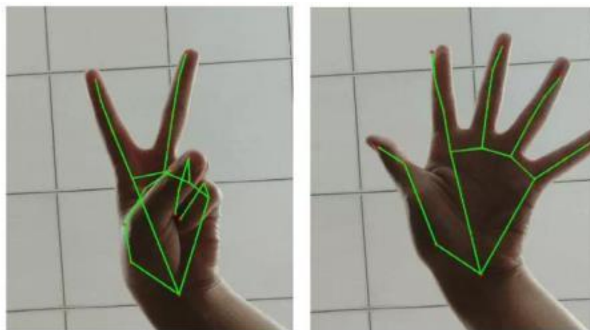


图 2-7 指关节跟踪示意图

□ 手势识别

手势识别是将模型参数空间里的轨迹 (或点)分类到该空间里某个子集的过程，其包括静态手势识别和动态手势识别，动态手势识别最终可转化为静态手势识别。从手势识别的技术实现来看，常见手势识别方法主要有：模板匹配法、神经网络法和隐马尔可夫模型法。

模板匹配法是将手势的动作看成是一个由静态手势图像所组成的序列，然后将待识别的手势模板序列与已知的手势模板序列进行比较，从而识别出手势。基于模板的方法不仅是指相互对应的模板匹配算法，也包括对时间、空间进行规整后的模板匹配算法，如动态时空规整算法和动态规划算法。由于动作的快慢不同，图像序列中的每幅图像较难做到相互对应，因此进行时间上的规整是模板匹配的重要步骤。代表性的模版匹配法是动态时间规整（Dynamic Time Warping, DTW）。

神经网络法：该方法具有分类特性及抗干扰性，具有自组织及自学习能力，具有分布性特点，能有效抗噪声并处理不完整模式，还具备模式推广能力。然而由于其处理时间序列的能力不强，目前广泛应用于静态手势的识别。

隐马尔可夫模型法是一种统计模型，用隐马尔可夫建模的系统具有双重随机过程，其包括状态转移和观察值输出的随机过程。其中状态转移的随机过程是隐性的，其通过观察序列的随机过程所表现。

AMiner

2.3.3姿势识别

姿势识别通俗的讲就是研究一种数据的分类问题，只不过这里的数据是时变数据，更简洁的说就是把需要进行测试的数据中的特征按照某种规则与实验之前定下的身体的一些典型姿势的数据的特征进行一一匹配。所以研究身体姿势识别的实质就是研究如何从样本中学习获取一组典型的身体姿势的参考序列。对身体姿势进行识别，常用的算法有三类：（1）基于模板匹配的身体姿势识别方法；（2）基于状态空间的身体姿势识别方法；（3）基于语义描述的身体姿势识别方法[15]。

- 基于模板匹配的身体姿势识别方法

基于模板匹配的姿势识别方法是将姿势序列转化为静态的模型来处理。大致思想是：首先需要获得一个姿势模板库，这就需要对身体姿势进行训练；然后计算需要测试的身体姿势的数据与模板库的姿势的数据进行匹配的相似度，让相似度最大的属于模板库所对应的姿势类别作为实验测试数据的最终结果。这种方法的研究重点是如何找到一个比较好的模板库。基于模板匹配的姿势识别方法的优点是操作和实现起来都比较简单方便，因为这类方法并不需要耗费大量的人力和财力，同时过程中涉及的计算很少。缺点是由于模板属于一种静态的方式，没有考虑到动态之间的一些关系，例如运动序列中相邻时序之间是有相互制约关系的，另外这种方法对噪声和运动时间间隔的鲁棒性都比较差。基于模板匹配的身体姿势识别方法所使用的姿势序列的静态模型主要有：运动能量图像、运动历史图像、二维网络、动态时间规整。

- 基于状态空间的身体姿势识别方法

状态空间的身体姿势识别方法的实质是将姿势转化为状态和联合概率。主要思想是：从运动的图像例如视频中，将图像序列进行分解得到一系列的静态姿势，或者直接把姿势序列看成是一个状态结点，这些静态姿势或者状态结点之间并不是完全独立的，由给定的联合概率关系将它们联系到一起。那么我们获取的全部的姿势序列就被当成是选取的静态姿势在这些特定状态结点之间的一个状态切

AMiner

换，通过计算每个状态到达其它状态的联合概率，选取最大的联合概率作为分类的准则。基于状态空间的身体姿势识别方法的优点是：可以很方便的对时间序列进行预测，缺点是算法本身的思想导致计算复杂度的增加。空间的身体姿势识别方法可以分为：隐马尔可夫及其改进模型、动态贝叶斯网络、人工神经网络、支持向量机。

- 基于语义描述的身体姿势识别方法

语义描述的身体姿势识别方法是一种利用指定的语法格式将场景中人物的活动情况用填空的形式填入到对应的语法格式中，其实质就是用自然语言达到对场景描述的目的。其语法格式很简单，其实就是语文中常见的语法，主语人、谓语人的动作、宾语实物，采用这种方法来对身体姿势进行识别的应用条件是这段时间范围不能间断。我们在用这个方法进行识别的道路上还只是处于起步阶段，后期还需要加强研究和加大力度应用，尤其要注意场景的复杂性等因素。基于语义描述的身体姿势识别方法主要是有限状态机和上下文无关文法。

2.4眼动交互

2.4.1发展现状

利用眼动跟踪技术记录人的眼球运动数据及其对应的视觉注意行为，获取用户当前的视觉注意焦点等时-空参数，对用户视觉感知和认知活动进行分析推理，进而为人机交互提供数据输入和控制输出。

为了开展基于眼动跟踪的人机交互研究，首先需要对眼动行为进行特征分析与参数化，常用的参数包括注视点数目、注视点持续时间、眼跳长度及其派生的其他各种度量参数。如图 2-8所示，在此基础上利用各种参数实现基于数据驱动的行为分析和交互控制，前者主要包括针对用户界面的可用性评估、眼动数据可视化与可视分析；后者主要包括眼动跟踪算法及相应眼动仪等装置的研发，以及具体的交互技术和应用。



图 2-8 基于眼动跟踪的人机交互研究框架

近年来，随着人工智能技术的发展，眼动跟踪作为感知和理解用户的关键环节，更加引起相关领域学者的关注。利用人工智能技术提高眼动计算的精度和效率，对人的感知和认知状态进行深入理解，构建“人在回路”的智能人机交互框架，实现用户主导的自动化系统、基于人机共生的 AI系统。下面介绍几个主要的研究工作：

□ 基于群智感知的眼动计算与分析

利用人的短时记忆特性建立注视点回忆和自我报告机制，通过众包技术在群智环境下实现用户注视点回忆任务的发布和大量数据的快速获取，进而基于大数据技术实现上下文感知的任务推荐和误差校正，有效提高注视点计算精度。在此基础上开展视觉显著性模型性能的高效优化，以及视觉设计和用户界面的在线评估等应用，不仅可以保证和实际眼动装置相当的计算精度，还能显著提高效率、降低软硬件和人力成本。

□ 基于大数据学习的眼动跟踪

传统的眼动跟踪算法主要基于眼部图像特征建立与坐标空间的映射模型，但这种算法依赖人工经验进行显式特征提取，且对外部光照条件、用户头部运动等外在因素敏感，计算复杂度高、稳定性低。为此，研究人员提出基于外观模型的计算方法，通过搜集大量不同用户及外在因素下的眼部图片，利用卷积神经网络等机器学习算法自动提取隐式特征，并建立与外在坐标空间之间的映射模型，虽

然计算精确度还有待提高，但计算过程简单、效率高，易于在互联网、移动端等环境进行轻量级的部署，实现面向电子阅读、手机游戏、广告推送等方面的便捷智能交互应用。

□ 眼动数据与脑电数据融合的智能交互

在交互任务执行过程中，记录由视觉注意行为引起的眼动和脑电数据，提取这些数据的特征并进行参数化。在此基础上，进行数据降噪处理和插补遗失数据，并挖掘视觉分布的模式和知识，生成高层的交互语义，以此对眼动和脑电数据进行时空同步处理和序列分析。通过训练人工神经网络，研究眼动和脑电数据与心理认知负荷之间的量化映射关系，进而依托隐式的自适应用户界面，降低用户认知负荷，提高交互效率，在网上课堂、汽车安全驾驶、医疗影像分析等方面提供智能辅助支撑，促进相关产业的创新实践。

2.4.2常见交互方式

眼动（Eye Movement），即眼球的运动。早期，研究者们发现人们阅读过程中的视线除了进行注视和平滑扫视外，还存在一系列短暂的跳动和停顿。这一发现使得人们开始进行眼动行为的相关研究和探索 [16]。在眼动交互过程中，需要基于基本的眼动形式，设计相应的眼动人机交互方式或策略，使视线具有一定的触发意图，从而执行对应的控制指令。常用的几种眼动交互方式分析如下：

□ 驻留时间触发

在注视行为中，注视点的持续时间被称为驻留时间。驻留时间触发是指当注视点的驻留时间达到一定程度后，可以利用视线代替鼠标点击或键盘按钮等传统输入设备，触发相应的执行操作。驻留时间触发多用于控制图形界面或定位鼠标光标等，是一种较为流行的眼动交互方式，它也能够反映用户有意识的控制意图，以更好地完成交互。

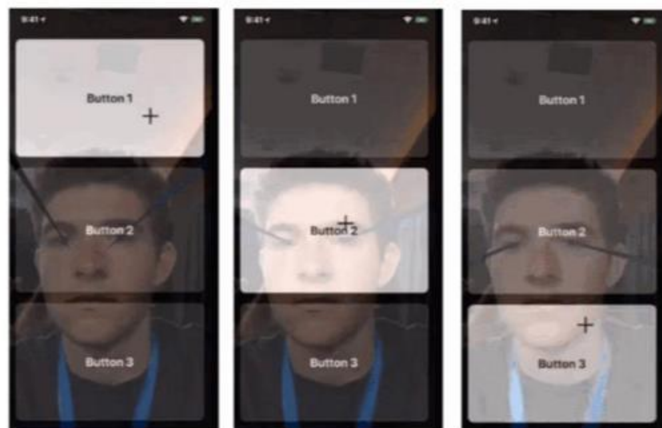


图 2-9 不同注视位置触发示意图

- 平滑追随运动

平滑追随运动多发生于观察场景中有缓慢移动的物体或目标，视线会产生平滑追随的运动状态。平滑追随运动是一种连续反馈的状态，眼睛捕捉运动目标的信号，将目标运动速度、方向、角度等信息反馈给大脑，再控制眼球跟随目标物体发生相对运动。在此过程中也会存在一些无意识眼跳等其他行为，在没有运动目标的场景下，一般不会产生该眼动行为，因此平滑追踪触发一般不是一种常用的眼动交互方式。

- 眨眼

眨眼是人在自然状态下的一种生理行为，是无意识的，正常情况下，普通人每一次眨眼的时长约为 200-400 ms，平均每分钟眨眼 10-20次。使用眨眼行为进行交互时，需要识别有意识的眨眼，例如眨眼频率超过一定程度，或一次眨眼过程中眼睛闭合的时间超过某个阈值。眨眼触发较为简单，但是当人眼处于长时间闭合状态时，由于眼动追踪仪无法捕捉瞳孔，可能会导致注视点的丢失，在一定程度上会影响眼控系统精度。

- 眼势

眼势是在眼跳的基础上提出的，但与眼跳的不同之处在于，眼跳往往是人在观察场景或对象时发生的一种无意识的视线转移，其眼跳的起点和终点都未知，依赖于人的视觉注意。而眼势被定义为一系列有序的视线行程，每一个行程是两

AMiner

个固定注视点或注视区域的有意的视线移动。因此，眼势作为一种新的眼动交互方式，可以反映人的有意识触发意图。不同路径的行程可以定义不同的眼势，不同的眼势可以映射为不同的交互指令。眼势可以分为单行程眼势和多行程眼势，如图 2-10所示。

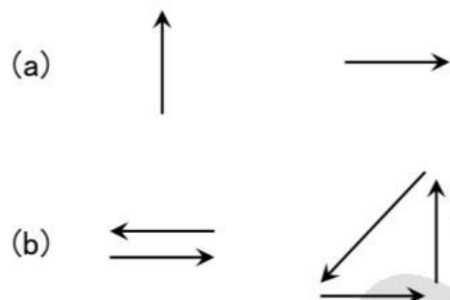


图 2-10 (a)单行程眼势；(b)多行程眼势

与其他眼动交互方式相比较，眼势具有以下几个特征：一是眼势能够将有意图的交互指令与无意识的眼跳行为区别开来，从而有效解决“米达斯触碰”等问题；二是与驻留时间触发相比，眼势控制区域不会占用较多的界面空间，甚至可以不依赖于界面；三是眼势的行程和起始位置都可以根据需求自己定义，具有一定的灵活性和多样性，且每个眼势可以对应一个交互指令，比较适用于指令多样化的交互场景。

2.5虚拟现实输入

近年来，虚拟现实技术（Virtual reality, VR）迅速普及，其产品已面向消费者市场，被广泛应用于娱乐、办公[17]、协作[18]、培训以及教育[19]等行业。文本输入作为应用中重要的交互技术，为应用提供了重要的交互体验。目前已经开发了多种适用于虚拟现实的文本输入技术，本节从输入机制角度回顾了现有的 VR 文本输入技术。

AMiner

2.5.1实体键盘技术

标准物理 QWERTY键盘是现有的高效舒适的桌面文本输入方式，如何将该键盘应用于虚拟现实环境已成为重要的研究点。移动场景中很难使用 QWERTY键盘进行输入，因此有研究使用了小型移动键盘进行 VR中的输入。

□ 基于 QWERTY实体键盘技术

基于 QWERTY实体键盘的文本输入技术利用了其使用广泛性以及用户熟悉度，降低用户学习成本。研究通过将实体键盘以及用户手部融合和虚拟现实环境，保证用户能够在虚拟现实头戴显示器（Head-Mounted Displays, HMDs）中正常输入，如图 2-11所示。



图 2-11 用户在虚拟现实中使用QWERTY实体键盘进行输入

早期研究直接将实体键盘引入虚拟现实中。James Walker 提出了 HMDs 中完全封闭（键盘不可见）情况下的基于 QWERTY键盘的 VelociTap辅助输入系统[20]，输入速度为 35个单词每分钟（35WPM），错误率 3.5%，并且在 2017年，他改进了该系统，加入了辅助输入的虚拟键盘以及解码器进行自动校正[21]，通过提供视觉反馈，输入速度达到 40WPM，错误率低于 5%。

通过跟踪人体手部位姿和键盘位置，将手部渲染到虚拟环境中，解决手部遮挡问题。Kim 等人通过最后一次按键位置将手部渲染到虚拟环境中，与使用摄像头相比该方式成本低但是输入精度较低[22]，具有 60%的桌面输入速度以及 80%的准确率。Bovet 等人使用 Logitech BRIDGE SDK 和 Logitech G810 Orion Spectrum键盘在 HTC Vive Pro HMD中渲染了 VR虚拟键盘以及手部，用户输入[23]

速度达到 44.4WPM，并且具有较小的精神和身体需求。Grubert[24] 认为这种方法可以保留至少 50％的用户桌面打字技能。但是，此方法的有效性取决于其跟踪传感器的可靠性。

另一种方式是直接将现实世界视频流直接嵌入虚拟环境中，将手和键盘的全部视图或部分视图呈现在虚拟环境中，帮助用户使用实体键盘。该方式通过对图像进行处理，将键盘和用户手部图像融合到环境中。 McGill[25] 提出了增强虚拟（AV）系统，该系统在绿幕背景中使用，输入速度达到 38.5WPM。Jiang[26] 等人结合深度相机和彩色相机进行实时视频处理，避免了光照不稳定问题，保留了 78%的桌面文本输入性能。

Grubert[24] 等人比较了 HMDs 中使用实体键盘和触摸屏的打字性能，实验发现新用户在使用桌面键盘时具有正常使用实体键盘时约 60％的打字速度，而使用多点触控虚拟键盘时具有约 40-45％的打字速度。

QWERTY 键盘仍然是虚拟环境中的基本输入方式，用户手部位置感知问题是该种输入技术需要解决的关键问题。

□ 基于移动实体的输入技术

一些沉浸式环境中空间有限，并且用户需要进行移动，QWERTY 实体键盘限制了此功能，有研究者提出了小型实体可移动键盘解决该问题。 Bowman[27] 等人和 González[28] 等人曾将 Twiddler用于 VR。Twiddler是一个具有 12键的弦键盘，通过按下一个键或者多个键的组合完成字符的输入，其输入速度为 3.0WPM，正确率 82%。

González等人还将 9键移动键盘用于 VR，每个键含有 3个字符，通过单次或者多次按键实现字符的输入，输入速度为 12.1WPM，正确率 95%。

□ 虚实混合技术的影响

实体键盘技术中虚拟混合技术会影响用户的输入表现，研究者比较了不同手部和键盘呈现方式对输入的影响。Grubert[29] 等人对比了 HMDs 虚拟环境中四种不同手部呈现方式（无虚拟手、三维手部模型、仅显示指尖、通过使用绿幕混合

视频显示真实手部图像）对文字输入的影响，实验发现在仅显示指尖和显示手部图像两种方式下使用 QWERTY 键盘时用户具有较好的表现。Knierim[30]等比较了四种情况下的不同手部表示：无，真实手，抽象手和指尖，发现手的呈现方式不会影响经验丰富的打字员的输入速度或准确性，但是影响了没有经验的人的进入速度。Lin[31]等人比较了虚拟环境中虚拟键盘是否存在以及手部和键盘与虚拟环境的融合方式对输入效率的影响，发现不同方式对于输入效率的影响并没有显著性差异。

2.5.2 虚拟键盘技术

虚拟键盘文本输入技术是目前常用的输入方式之一，解决了对实体键盘的依赖性问题，但目前的虚拟键盘仍然保持着 QWERTY键盘的布局模式，利用用户熟悉度，降低用户学习成本。虚拟键盘的输入也利用了不同的确认方式，包括手柄射线，头部射线，手部、手指、手腕手势，触碰眼动，手写等方式。

□ 基于手持控制器的输入技术

手持控制器作为常用的交互设备，也是重要的文本输入设备。基于手持控制器射线的虚拟键盘输入技术是目前商用 HMDs 最常用的输入方式，通过带有运动跟踪器的手柄发出射线指向虚拟键盘，按键进行确认选择，完成文本输入。

Speicher[32] 等人比较了六种基于虚拟 QWERTY键盘的输入技术：头部指向，手持控制器指向，手持控制器作为类手写笔，手指点击，手持控制器作为操作杆使用光标进行连续以及离散的输入。其中，手持控制器的指向输入方式在输入速率和错误率方面具有较好的表现，达到 15.44WPM和 0.97%的错误率；手持控制器作为类手写笔进行点击输入，输入速率为 12.69%WPM，错误率 1.94%；手持控制器作为光标进项离散输入的输入速度为 5.31 WPM，错误率 2.79%，进行连续输入时，输入速度 8.35WPM，错误率 2.15%。手柄控制器用于输入具有良好的输入表现，但是长时间使用会造成用户身体疲劳。Min[33] 等设计了一个需要消歧的 QWERTY 键盘，该键盘将按键排列成 3×3 的网格，用户首先通过按一个按钮选择一个单元格，然后通过一次或多次按该按钮选择目标字符，分别对应于

AMiner

该字符在单元格中的位置，这种方法由于尺寸较小而节省了空间，为工作留出了额外的空间，但该方法尚未评估。

Chen[34]等人使用手持控制器射线在虚拟键盘上进行单词级别的文本输入，并与使用压敏触摸屏设备进行单词手势输入进行了比较，其中手持控制器的输入速度可达到 34.2 WPM。

Xu[35]等人比较了四种虚拟现实中的文本输入技术：手持控制器作为光标、头部射线暂留、手部位置和手势选择、头部射线和手势选择，并基于四种技术比较了滑动和点击两种不同的输入方式，其中，基于手持控制器的输入方式具有较好的输入表现，滑动输入速率为 13.68WPM，点击的输入速度为 14.6WPM。

□ 基于头部的输入技术

头部指向技术也是 HMDs中默认的交互方式之一，该方式释放了用户双手。该技术类似手持控制器的文本输入方式，通过头部射线进行指向，结合暂留技术或者按下控制器的键值来进行选择。

在 Speicher[32] 的对比研究中，头部射线指向技术输入速度为 10.20WPM，错误率 1.15%。此外，该方式具有较好的用户体验，用户具有较小的挫败感。Yu[36]等人通过头部射线指向虚拟键盘，使用按键点击和暂留技术实现文本输入，并与手势输入技术进行了比较。其中，控制器按键点击方式输入速度为 15.6WPM，错误率 98%，而选择 Majaranta[37] 建议的 400ms暂留技术的输入方式输入速度为 10.6WPM，错误率 4.2%[35]。Xu 等人的研究中基于头部射线的暂留技术使用 500ms 的暂留时间进行选择确认，该种输入方式在四种输入方式中具有最差的表现，输入速度为 5.62WPM，其中滑动输入速度为 7.94WPM；头部射线与手势结合的输入方式中，点击的输入速度约为 10WPM，滑动的输入速度约为 8WPM。暂留技术可能会影响输入表现。

基于头部的输入方式需要头部不断的运动，会造成疲劳以及晕动症。该种方式比较适合于少量文本，例如密码的输入。

□ 基于手部的输入技术

通过使用外部摄像机、传感器等可以跟踪用户手部运动，很多研究探索了该种基于手指、手腕和手势的输入技术。

Bowman[38]使用数据手套跟踪用户手部，将 QWERTY 键盘映射到手指。例如，将第二行字母“a”，“s”，“d”和“f”分别映射到左侧的小指，无名指，中指和食指。通过这种方法，用户首先旋转手以选择一行，然后通过拇指和手指的捏合动作来输入字符。 KITTY 键盘使用类似的方法，但是它没有使用手旋转，而是使用具有六个自由度的[39]拇指来选择一行。拇指前面的三个位置和拇指后面的三个位置分配给不同的行。例如，在左手小指和中间拇指中间捏一下输入字母“a”。DigiTouch 通过跟踪手指上的电阻式织物界面上的压力检测用户点击位置[40]，进行对应文字的输入。FingerT9[41]则将手指分为不同区域，将传统 T9 键盘映射到手指，通过检测点击实现文本的输入。

TipText 将 QWERTY键盘映射到食指指尖（如图 2-12（a）所示），并将虚拟键盘划分[42]为不同的区域，但保持相同的字母排序，实现单词级别的输入，同样的，BiTipText 则将键盘分为两个部分分别映射到双手的食指指尖（如 2-12（b）所示）[43]，实现单词级别的输入。

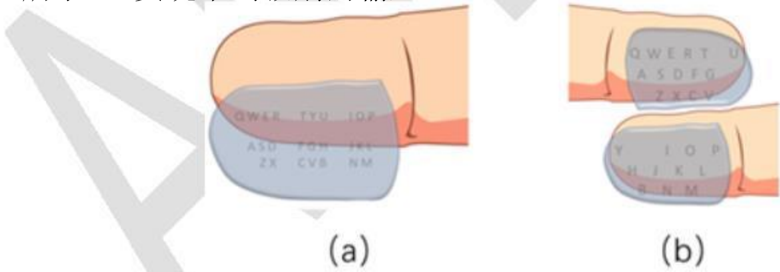


图 2-12 (a) TipText；(b) BiTipText

RotoSwype[44]通过在食指佩戴传感器，结合虚拟 QWERTY键盘，通过手势实现虚拟现实中单词级别的输入，输入速度达到 14.8WPM，错误率 0.5%。

□ 基于触摸屏的输入技术

研究者要利用用户对触摸屏设备的熟悉度、肌肉记忆等将触屏设备引入虚拟现实进行交互。

Gugenheimer 等人在 HMD 上增加了 17.78 厘米的电容式触摸板，用户能够通过使用触摸板[45]在浮动虚拟键盘上选择字符来输入文本，在非正式评估中，此方法达到了 10WPM，该方法限制了用户交互界面在 HMD背面，长时间的使用也将导致手部疲劳。Kim[46]通过使用三星 Galaxy S4智能手机进行文本输入，通过强电容屏可以感知和触摸屏具有一定距离的手指，并在虚拟键盘显示手指的位置，帮助用户进行输入，并加入矫正算法提升输入效率，输入速度为 7.8WPM和 9.0 WPM，准确率分别为 79.5%和 92.6%。Lu 等人开发了 BlindType探索虚拟现实中的无眼输入，通过利用用户拇指肌肉记忆[47]在手持设备进项文本输入，并加入解码算法，输入速度为 22.77WPM，正确率为 98.29%。

 基于视点的输入技术

通过在 HMD设备上配置眼动设备，可以通过眼动进行输入。Hajana 研究了 VR 在平面和曲面虚拟键盘中的注视打字[48]，用户可以通过在虚拟按键上停留 520毫秒或按下控制器按键来选择字符。其中，驻留技术的输入速度为 9.4WPM，控制器确认的输入速度为 7.5WPM，控制器确认方式在平坦和弯曲键盘输入速度分别为 10.2WPM和 9.2 WPM。两者的准确率均超过 99%。在弯曲键盘和平面键盘之间未发现明显差异。 Yao[49]等人结合了带有注视功能的大脑计算机接口（BCI），可以在 VR中输入文本。他们将来自大脑的电信号与注视结合起来，以确定光标的位置并进行选择。该方法未使用 QWERTY，而是设计了一个具有多行且每行 8个字符的字母布局（即，第一行包括从“a”到“h”的字母）。非正式研究中，其输入速度为 10WPM。

基于视点的输入方式受视点检测的精度的限制，以及人眼本身的眼动等都会影响输入的精度。但是该方式比较适合肢体残疾的用户。

 基于空中敲击的输入

虚拟 QWERTY键盘可以直接渲染到虚拟环境中，用户模拟真实环境敲击真实键盘进项输入，具有直观性，降低用户学习成本，但缺乏力反馈会导致用户手部疲劳，此外，点击动作的判断准确率也会影响文本输入效果。

Hsieh[50]等人通过数据手套跟踪用户拇指、食指、中指，通过跟踪手部位置确定手部对应的虚拟键盘的位置，通过不同手指的点击动作实现虚拟环境中的文本输入，但并未进行评估。TiTAN 使用深度相机检测用户空中点击动作，允许用户在半空中通过模仿他们通常在物理键盘或触摸屏上的打字动作来输入文本，并比较了使用两个手指、十个手指以及书写的用户表现，其中，两手指和十手指的输入速度分别为 13.76WPM、13.57WPM，错误率分别为 7.3%和 9.4%。

Dudley[51] 等人提出 VISAR键盘（如图 2-13所示），支持用户使用单指头通过虚拟输入表面进行文本输入，经过优化，输入速度为 16.76WPM，错误率 0.63%。Dudley 探究了虚拟键盘敲击在是否存在物理表面以及使用不同手指数的表现，实验结果表明，在具有物理界面的桌面上敲击具有较高的用户满意度，在使用无物理界面支持的虚拟键盘时，用户更喜欢使用双指进行输入而不是十指。其中，在具有物理界面支持的双指输入的输入速度为 55.6WPM，十指的输入速度为 51.6WPM，无物理界面的空中点击双指输入速度为 42.1WPM，十指输入速度为 34.5WPM。



图 2-13 VISAR键盘

2.5.3新型输入技术

□ 手部输入技术

研究者提出了基于手部的文本输入方式，通过利用手部触觉，结合其他传感器、输入技术来进项文本输入。

Kim[52]等人通过在食指上佩戴小型触摸表面，将字母分配在 2×3 的网格中（如图 2-14 所示），通过拇指点击表面不同区域，使用两步输入技术，实现单手无眼输入，输入速度 11.41WPM，总错误率 11.96%。Jiang[53] 等人通过将字符

AMiner

分配到食指和中指的六个区域内，结合两步输入法，通过拇指点击动作在 VR中进行单手输入，输入速度 9.82WPM，错误率 6.03%。该方法通过超薄压传感器进行点击检测。BlueTap[54]键盘按字母顺序将字母映射到手指关节上，每个手指最多四个字符。用户用拇指点击手指的不同部分以选择字符，使用腕戴式摄像机来检测点击。Pratorious[55]等将 T9键盘映射到手指关节，每个指关节或指尖最多可包含 4个字母，使用类似 T9键盘的多次点击输入方式进行文本输入，该方式利用了腕戴式摄像机和加速度计来检测点击，输入速度为 10WPM。Ogitani[56]等人设计了一种 12键移动键盘布局，该布局最多可将 4个字符映射到每个键。在这种布局下，用户点击包含目标字符的键，然后向字符方向滑动以输入字符，并使用键盘评估了两种技术。第一种技术将键盘投影在手掌上，用户使用另一只手的食指选择字符。第二种技术在空中显示键盘，用户可以用手来选择字符。输入速度分别为 5.6 WPM和 8.2WPM。



图 2-14 小型触摸表面

□ 圆形键盘输入技术

Yu[57]等人开发了 PizzaText，将一个圆圈分成 7个切片，每个切片包含 4个字符（如图 2-15（a）所示）。用户使用操纵杆的双指拨杆与该布局交互-右指拨杆用于在圆形键盘周围移动，而左指拨杆用于选择字符，输入速度为 15.9 WPM和准确率 94.6%，并评估了该键盘布局的三种不同的尺寸，但是，三种布局之间输入速度和准确率并未有显着差异。 Xu[58]等人通过结合头部视点和圆形键盘技术 RingText（如图 2-15（b）所示），通过用户的头部移动控制的虚拟光标来进行选择-进入字母区域会触发选择，而移回内部圆圈会重置选择，输入速度为 13.24WPM，准确率 97.56%。

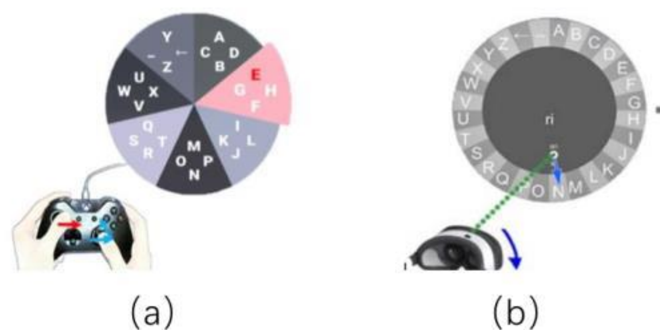


图 2-15 (a)PizzaText；(b)RingText

□ 立体输入技术

3D物体现实是 VR特点之一，但目前的大多数用于 VR的键盘都是 2D的。立方键盘[59]是 3D键盘，其将字母排列成 3×3×3的 3D阵列，包含 27个单元格，其中 26个代表 26个英语字母，中间是一个空白单元格。用户使用控制器在单元格中导航以选择字符，该方法输入速度为 21.7WPM。3D布局的 VR键盘仍需要更多的探索。

目前尚未有一种通用的高效的适合于虚拟现实的文本输入技术，目前的输入技术都是基于现有技术迁移、优化而来。虚拟现实具有不同的应用场景，不同场景会对输入技术提出更具体的要求，例如虚拟办公场景需要保证快速准确的输入以及便携性，针对不同使用场景的文本输入技术人仍需要等多的探索，包括对现有技术的优化以及新的技术的开发等等。

2.6多模态交互

自从 Bolt 发表了开创性的论文 *Put that there:Voice and Gesture at the Graphics Interface*，可用于计算机应用程序交互的多模态输入就成为了人机交互研究的一个活跃领域。这种不同形式的输入组合（例如，语音、手势、触摸、凝视等）被称为多模态交互模式，其目标是向用户提供与计算机进行交互的多种选择方式，以支持自然的用户选择。

相比于传统的单一界面，多模态界面可以被定义为多个输入模态的组合，这些组合可以分为 6种基本类型：

AMiner

互补型：当两个或多个输入模态联合发布一个命令时，它们便会相得益彰。例如，为了实例化一个虚拟形象，用户做出指示手势，然后说话。语音和手势相得益彰，因为手势提供了在哪里放置对象的信息，而语音命令则提供了放置什么类型的对象的信息。

重复型：当两个或多个输入模态同时向某个应用程序发送信息时，它们的输入模态是冗余的。通过让每个模态发出相同的命令，多重的信息可以帮助解决识别错误的问题，并加强系统需要执行的操作。例如，用户发出一个语音命令来创建一个可视化工具，同时也做一个手势表示该工具的创建。当提供多于一个的输入流时，该系统便有更好的机会来识别用户的预期行为。

等价型：当用户具有使用多个模态的选择时，两个或多个输入模态是等价的。例如，用户可以通过发出一个语音命令，或从一个虚拟的调色板中选择对象来创建一个虚拟对象。这两种模态呈现的是等效的交互，且最终的结果是相同的。用户也可以根据自己偏好（他们只喜欢在虚拟调色板上使用语音输入）或规避（语音识别不够准确，因此他们改用调色板）来选择使用的方式。

专业型：当某一个模态总是用于一个特定的任务时它就成了专业的模态，因为它是比较合适该任务的，或者说对于该任务来说它是当仁不让的。例如，用户希望在虚拟环境中创建和放置一个对象。对于这个特定的任务，做出一个指向的手势确定物体的位置是极具意义的，因为对于放置物体可能使用的语音命令范围太广，并且一个语音命令无法达到对象放置任务的特定性。

并发型：当两个或多个以上的输入模态在同一时间发出不同的命令时，它们是并发的。例如，用户在虚拟环境用手势来导航，与此同时，使用语音命令在该环境中询问关于对象的问题。并发型让用户可以发出命令并执行命令，其体现为在做晚餐的同时也可也以打电话的真实世界的任务。

转化型：当两个输入模态分别从对方获取到信息时它们就会将信息转化，并使用此信息来完成一个给定的任务。多模态交互转化的最佳例子之一是在一键通话界面里，语音模态从一个手势动作获得信息，告诉它应激活通话[60]。

AMiner

2.7信息无障碍中的智能交互技术

信息无障碍（information accessibility）是一个学科交叉的技术和应用领域，旨在用信息技术弥补残障人士生理和认知能力的不足，让他们可以顺畅地与他人、物理世界和信息设备进行交互。据中国残联统计，中国现有 8500万残疾人，是世界上残疾人口最多的国家。其中，听力残疾 2000万人，视力残疾 1200万人，各类肢体残疾 2500万人，智力残疾和精神残疾 1200万人。随着社会老龄化程度加重，残疾人口数量也在持续增长。互联网和用户终端的普及，使得信息无障碍成为一个越来越值得关注的领域，目标是解决残障人士的信息访问甚至是生活服务问题。

信息无障碍始于个人计算设备，IBM在 1984年首次开发了基于桌面操作系统的读屏程序，让盲人可以使用电脑。1997 年，万维网联盟（W3C）成立了网络无障碍推动(WAI)小组，推动网络页面的信息无障碍。随着人工智能技术的发展，语音识别、图像识别、手语翻译等技术也被应用到信息无障碍领域中，支持更多的残疾用户（比如聋哑人）通信和访问信息设备。微软在 2018年推出了“人工智能无障碍计划”（AI for Accessibility），国内的互联网公司阿里巴巴、百度等也积极推出智能读图等无障碍交互应用。

2017 年，人机交互领域的旗舰会议 ACM CHI（ACM CHI Conference on Human Factorsin Computing Systems）将无障碍列为十大方向之一。我国于 2004 年举办了第一届中国信息无障碍论坛。2006 年，工信部把信息无障碍工作纳入“阳光绿色工程”，并由中国通信标准化协会（CCSA）开始系统化制定信息无障碍标准。2008 年，工信部发布了首个网站设计无障碍技术标准，信息无障碍在“十三五”规划期间被纳入国家发展计划。从研究和应用水平上看，信息无障碍总体还处于比较初步的状态。在应用上，针对信息访问和设备使用，具有基本功能的技术可以被应用，但效果和效率等可用性指标都不高；在现实生活中，针对听障人士与他人交流、盲人独立出行等，能支撑的新技术还处于原型和概念阶段。本节将对无障碍交互领域的现状、科学问题及相关技术进行简要介绍[61]。

2.7.1信息无障碍交互技术的现状

残障类型多样，所需要的无障碍技术也不尽相同，这里介绍三类主要的残障类型（视觉障碍、听觉障碍和运动障碍）人群遇到的问题和主要的技术解决方案。

视力残疾用户的需求包括独立出行、识别身边物体、与信息设备交互等。针对独立出行的需求，目前有基于计算机视觉的道路识别技术，通过立体声场或者震动反馈为视力残疾用户指示方向。但是这些设备目前还不能取代盲杖，还需要更多的技术突破。针对识别物体的需求，主要是利用视频 /图像转换为文本的技术，包括微软的 Seeing AI和谷歌的 Lookout都是此类应用。针对使用手机和电脑的需求，主要采用读屏程序 Screen Reader（苹果手机上的 VoiceOver或者安卓系统上的 Talkback，均为系统默认自带功能），可以通过语音读出获得焦点的控件信息，这样视力残疾用户通过听就能了解设备界面上的信息内容。

听力残疾用户面临的主要问题是与人交流存在障碍，以及观看视频内容时听不到声音。老年听力障碍是指随着年龄增长，听觉器官的衰老和退变所导致的听觉功能下降，发病率居世界第三位。助听器设备通过放大声音信号，可解决“听不到”的问题；但对于听觉中枢受损的人，声音信号分析能力却难以弥补，解决不了“听得清”的问题。针对听障用户，相关信息无障碍技术包括手语的识别与合成，以及语音识别技术。

运动控制能力缺失的用户，包括上肢残疾，或者患帕金森症、脑瘫、肌肉萎缩、渐冻症等疾病的用户。他们丧失了灵活控制手指运动的能力，而手指是人表达交互意图的主要运动器官，也是电脑和手机的主要操作器官。在构建面向这类用户的信息无障碍交互技术时，其中一个难题是用户的差异性，几乎每个用户的可运动部位及其运动能力都是不同的，给构建适合于个体的通用输入技术带来了挑战。相关的技术方案有眼动，但是用“眼动”作为输入方式时，缺少“确认”操作，容易产生误触发，且操作精度有限。

2.7.2信息无障碍的主要科学问题

- 音视频的理解和信息转换

主要针对听障和视障。视觉和听觉是人们接受信息的主要感官。听障和视障用户因为缺乏某种感官而无法完整理解信息，需要建立音视频的理解技术，用机器算法理解音视频内容的语义，进而转换为用户可用感官能接受的信息类型，包括音频和文字之间的语音识别和文本到语言（Text To Speech, TTS）技术，图像到文字和视频到文字的技术。目前，精度是主要问题，尤其是克服多种噪声条件下的高精度实现，对于这些技术的可用性起到关键作用。

□ 图形用户界面到声音界面的编码转换

主要针对视障。个人电脑和手机都是图形用户界面，信息以可视的方式传递给用户，而视障用户只能通过听觉（触觉为辅）来接收信息，相比于视觉，不仅信息接收的带宽要低很多，而且信息呈现的模式也发生了变化。视觉提供整体和并行的信息获取能力，听觉只能提供局部串行的信息。这也会影响用户对于交互界面的心理模型，进而影响到交互决策。因此，需要研究从图形界面到声音界面的编码转换方法，优化“读屏”的方法。

□ 个性化信息输入和意图理解

主要针对视障和运动障碍。人体的运动控制系统包括运动执行和反馈两部分。运动障碍用户无法精确灵活地控制手指运动，视障用户由于缺少视觉反馈也不能做精确的输入控制，导致物理运动自由度受限和运动控制精度低的问题。前者需要开发具有个性化能力的输入技术，根据用户实际可以控制的输入范围来映射有效的输入；后者需要实现从有噪声的运动控制数据中提取用户的交互意图。

2.7.3通过智能交互技术实现信息无障碍

信息无障碍是以用户为中心的交互方案，是对人的交互性能的优化。优秀的信息无障碍技术要适应用户的生理和认知能力，而不是让用户适应技术。为此，要采用智能交互方法来开展研究，从用户角度来设计和创新适用的交互模式，通过智能传感、智能用户意图推理和智能信息呈现来构建信息无障碍的交互界面。

□ 对用户行为和认知能力的准确建模

AMiner

需要建立用户动作能力和心理模型的计算模型，建立并引入生理、心理的先验知识来描述用户的信息输入输出能力，对用户意图、表达方式、动作控制能力做统计建模和描述。残障用户的一个重要特点就是个性化，每个用户的信息输入输出能力都不同，除了需要研究合适的模型函数，还要研究个性化参数的计算方法，力求能准确地描述个体残障用户的能力。

□ 智能的感知技术

需要研究高精度的感知技术，感知用户的外在动作，也感知用户的内在心理状态。信息无障碍中，智能感知的应用场景非常丰富，针对不同类型的残障用户、不同的交互任务、不同的使用情景，都需要适合的传感方案。哪些动作或者心理状态是有交互价值的，如何采用成本可控、易于部署的硬件方案都是需要考虑的问题。

□ 智能的意图推理技术

残障用户通过自然动作（手、眼或身体其他部位）表达交互意图。优化信息无障碍的交互体验，需要使用户在表达意图时的生理和心理开销最小化，但这样将导致用户的表达方式和表达动作都是不精确的，在时间和空间上都存在随机性。如何从连续随机的行为数据中提取用户的输入意图，是需要解决的问题。作为操控型的交互界面，需要具有高精度、可理解、结果可预测的特点。

□ 智能的信息呈现

感官残障用户对于交互界面的理解是不完整的，交互决策的心理模型带有随机性。高可用的信息无障碍交互界面，首先需要有对用户信息需求的预测能力，确定信息输出的目标；然后根据用户的信息接收能力将目标信息编码到具体模态上，编码方式涉及多模态融合；最后根据交互情景，对信息呈现的编码方式做动态优化调整，保证用户接收信息的有效性。

AMiner

2.8人机交互领域必读论文

基于 AMiner 系统的“Topic 必读论文”功能，通过本领域热心专业读者推荐，本部分选取其中代表性的八篇论文进行解读。如欲查看完整人机交互领域必读论文，请查看网站：https://aminer.cn/search/pub?q=Human%20Computer%20Interaction&from=trreport。

1. iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing

论文链接：https://www.aminer.cn/pub/5550447a45ce0a409eb4e21d/

作者：Martin Weigel，Tong Lu，Gilles Bailly，Antti Oulasvirta，Carmel Majidi，Jürgen Steimle

作者单位：Max Planck信息研究所，萨尔兰大学，卡耐基梅隆大学，阿尔托大学等

该文荣获了 CHI 2015年最佳论文奖。作者在文中提出了一种可穿戴在皮肤上的传感器 iSkin，它使用与人体相兼容的硅胶制作而成，可设计成各种不同的样式和大小。它能够穿戴在人体的不同位置，包括手指、前臂，甚至是耳垂的后方。iSkin 的最大特点是薄，柔软并且能折叠弯曲，即使发生了严重的形变也不会影响到它的功能。它内嵌电容式和电阻式传感器，支持多点触控功能，可以识别划动等动作。这项研究使得新型的体内装置成为可能，包括手指穿戴设备、对传统可穿戴设备的扩展以及触摸式智能贴纸。作者希望他们的研究可以为未来的人机交互场景提供新的方向。



Figure 1. iSkin is a thin, flexible, stretchable and visually customizable touch sensor that can be worn directly on the skin. We present three novel classes of on-body devices based on iSkin: (a) *FingerStrap*, exemplified here with a strap on the index finger for fast, one-handed control of incoming calls; (b) *Extensions for wearable devices*, exemplified here with a rollout keyboard attached to a smart watch; and *SkinStickers*, exemplified here with (c) an input surface for a music player attached to the forearm, (d) a click wheel on the back of the hand and (e) a headset control behind the ear.

AMiner

2. Project Jacquard: Interactive Digital Textiles at Scale

论文链接：https://www.aminer.cn/pub/57d063bdac4436735428f4e7/

作者： Ivan Poupyrev，Nan-Wei Gong，Shiho Fukuhara，Mustafa Emre Karagozler，Carsten Schwesig，Karen E. Robinson

作者单位：Google

这篇文章获得了 CHI 2016 年最佳论文奖。这款由谷歌发明的神奇布料 Project Jacquard（提花织物），核心技术是由传导线编织而成的布料能够作为触控屏使用。文中提出的这种新颖的交互式纺织布料，利用大家熟知的纺织原料，包括棉、麻、蚕丝、羊毛纤维与导电纤维编成纱线，之后这些具有优越拉伸性能的纱线被织进布料中，整个过程只要稍微改一改普通织布机的程序，就能轻松做到，并不需要额外的工序。面料的结构、颜色、图案也可以根据设计师的要求，随意改变。为了与智能手机等设备互联，Project Jacquard包含了可传导信号、导电并且可以和各种材质的纱线纺织在一起的特殊纱线。此外嵌入到衣服当中的还有像纽扣一样大小的电路线圈，以及带芯片的独立主板；Project Jacquard可以用来捕捉人体的触摸输入以及手势操作。文中还描述了纱线、纺织品、服装和用户交互性的发展，并提出了可穿戴的交互式布料目前存在的机遇和挑战。



Figure 1. Project Jacquard envisions *seamless and fluid integration* of interactivity woven into everyday objects and environments.

3. Motivation as a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale

论文链接：https://www.aminer.cn/pub/555042a345ce0a409eb43c63/

作者：René F. Kizilcec，Emily Schneider

作者单位：斯坦福大学

为了解决哪些动机可以预测大规模开放式在线课程中学习者的行为，以及它们的预测能力如何，该文预估了每个入学意向预测的 10个行为结果指标的变化。开放的在线学习环境吸引了不同动机的学习者，为了系统地描述这些学习者的学习动机，作者开发了在线学习入学意向量表（ OLEI)，这是一个从开放式回答中获得的 13 项问卷，以获取学习者的真实观点。尽管学习者参加不同课程的动机不同，但作者发现每个动机都预测了学习者的主要行为结果。根据学习者的动机和行为模式，作者推断出他们通过参与课程来满足的各种需求，比如认识新朋友、学习英语。为了满足这些需求，作者提出了多种设计方向，包括任何特定课程外部的虚拟社交空间、改进对本地学习群体的支持以及模块化以促进课程内容的可访问性和组织能力。因此，动机为理解在线学习者和设计在线课程更好地支持学习者的需求提供了新的视角。

4. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations

论文链接：https://www.aminer.cn/pub/5550447f45ce0a409eb4e4df/

作者：Matthew Kay，Cynthia Matuszek，Sean A. Munson

作者单位：华盛顿大学，马里兰大学巴尔的摩分校

本文荣获了 CHI 2015年最佳论文奖。该文发现，对职业的图像搜索结果一定程度上夸大了性别的刻板印象。信息环境可以影响人们的感知和行为。在这篇文章中，作者的研究结果描述了性别偏见存在于各种职业的图像搜索结果中。作者实验性地评估了图像搜索结果中的偏见对人们选择代表哪些职业的图像以及人们对每个职业中男性和女性的普遍性看法的影响。他们发现在搜索结果中既有刻板印象的夸张，也有系统的女性代表性不足的证据。他们还发现，当搜索结果符合职业定型观念时，人们对搜索结果的评价更高，而改变图像搜索结果中的性别代表性可以改变人们对现实世界的看法。

5. Voice Interfaces in Everyday Life

论文链接：https://www.aminer.cn/pub/5aed146117c44a4438152a2f

作者：Martin Porcheron,Joel E. Fischer,Stuart Reeves,Sarah Sharples

作者单位：诺丁汉大学

该文荣获了 CHI 2018年最佳论文奖。语音用户界面（VUI）变得无处不在，它既可以通过智能手机嵌入日常移动应用中，也可以通过“辅助”设备嵌入到家庭生活中。然而，用户如何将这些设备运用在日常的社交互动中，仍然有待探索。通过收集和研究参与者在家中为期一个月使用 Amazon Echo 的音频数据，作者的研究记录了 VUI 用户的使用实践，以及如何在用户复杂的社交生活中实现这种使用。作者提供的数据显示了如何将设备嵌入到诸如家庭晚餐之类的对话环境中，同时实现各种活动。他们还讨论了 VUI如何很好地协调与对话的顺序组织。最后，作者给出了 VUI交互、请求和响应设计的问责制的含义，并对设计“会话”界面提出了概念上的挑战。

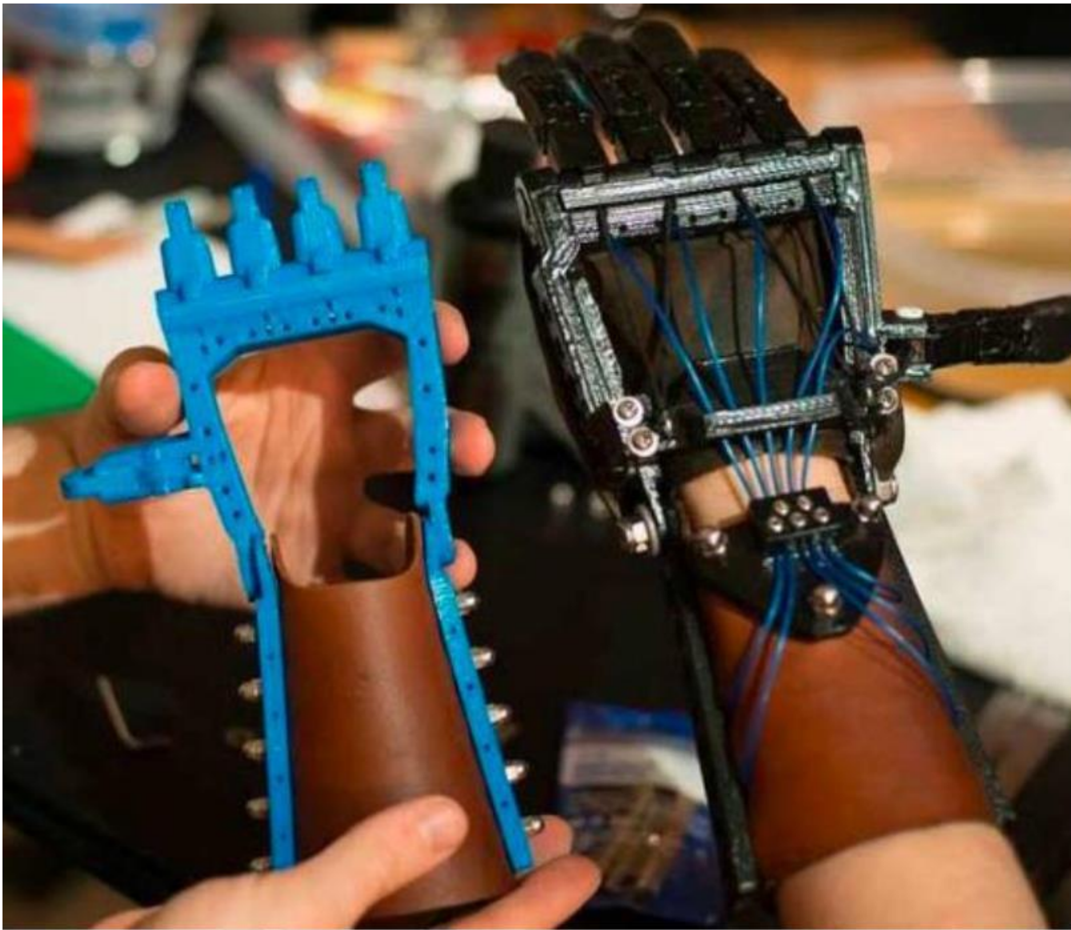


6. Sharing is Caring: Assistive Technology Designs on Thingiverse

论文链接：https://www.aminer.cn/pub/5550447a45ce0a409eb4e1c9

作者：Erin Buehler，Stacy M. Branham，Abdullah Ali，Jeremy J. Chang，Megan Kelly Hofmann，Amy Hurst，Shaun K. Kane

作者单位：马里兰大学，科罗拉多州立大学，科罗拉多大学



该文荣获了 CHI 2015年最佳论文奖。越来越多的在线社区支持开放源代码共享设计，可以使用快速原型技术构建物理对象。在这篇文章中，作者考察了在Thingiverse.com上发现的辅助技术的设计和动机，Thingiverse.com是撰写本文时这些社区中最大的一个。他们提供了 2008年以来发布到 Thingiverse的所有辅助技术的调查结果，并向设计师分发了一份调查问卷，探讨他们与辅助技术的关系以及创作这些设计的动机。这些设计中的大多数是计划在 3D打印机上制造的，包括为残疾人、老年人和药物管理人员提供的辅助设备和修改。这些设计中的许多都是由用户自己或代表朋友和爱人创作的。这些设计师在辅助技术的创造方面往往没有正规的培训或专业知识。本文讨论了该社区的趋势以及未来的机遇和挑战。

7. Acoustruments: Passive, Acoustically-Driven, Interactive Controls for Handheld Devices

论文链接：https://www.aminer.cn/pub/5550447c45ce0a409eb4e30a

作者：Gierad Laput，Eric Brockmeyer，Moshe Mahler，Scott E. Hudson，Chris Harrison

作者单位：卡内基梅隆大学，匹兹堡迪斯尼研究所

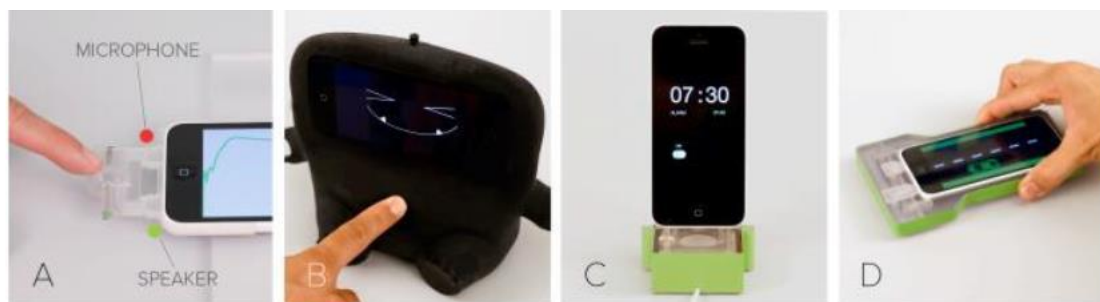


Figure 1. Like musical instruments, we introduce structural elements along the speaker-microphone pathway to characteristically alter the acoustic output. Using an expansive vocabulary of design primitives, we construct physical mechanisms (A), and on top of that, build end-user applications such as an interactive doll (B), an alarm clock (C), or an interactive toy car (D).

"声学仪器"（Acoustrument），研究人员利用低成本的 3D 打印塑料组件取代复杂的模拟电子电路，为智能手机打造出以声学驱动的多款创新外挂配件，如扩充埠以及其他有趣的扩充设备。在这篇文章中，研究人员描述如何设计出一些可沿着"扬声器—麦克风"路径放置的结构元素，从而改变声学输出特性。这一过程塑造出各种不同的实体机制，包括小型管径、可变形共振腔、提供多种选择音频路径的旋钮，以及更实际的将这些机制全部转换成各种不同的功能，例如旋钮、滑杆、近接与压力传感器、旋转编码器或甚至是倾斜传感器。这可为低廉的塑料App玩具扩展许多新应用范围，使其得以提供更丰富、有形且互动的功能。实验显示，这种声学仪器能够达到 99%的准确度，不但只需要一点点的练习或训练即可，而且它又十分抗噪声。

8. Interactivity Improves Usability of Geographic Maps for Visually Impaired People

论文链接：https://www.aminer.cn/pub/5550400d45ce0a409eb31e0c

作者：Anke Brock,Philippe Truillet,Bernard Oriola,Delphine Picard,Christophe Jouffrais

作者单位：艾克斯－马赛大学等

在这篇文章中，作者展示了交互式地图优于带有盲文图例的触觉纸质地图，它涉及三个可用性维度中的两个，并与第三个维度等效：效率，根据学习不同类型空间信息所需的时间来衡量。视觉障碍者使用触觉地图来获取空间的心理表征，但它们存在严重的局限性（信息量有限、盲文文本等）。这项研究提出了一个可用性的经典凸起线地图与互动地图组成的多点触摸屏、凸起线覆盖和音频输出。这两张地图都由 24 名盲人参与者进行了测试。研究结果显示，交互式地图是改善视力受损人群地图探索和认知地图绘制的一个很好的解决方案。以简单的听觉触觉互动取代盲文，可显著提高效率及使用者的满意度。

FIGURE 1. Photograph of a user exploring an interactive map.



Note. The raised-line map overlay is attached on top of the touch screen. The user is wearing mittens to prevent unintended touch input from the palms.



3人才篇

本篇通过 AMiner大数据平台对人机交互领域的顶级学术会议及期刊论文进行挖掘，获得这些会议和期刊最近 10 年的论文，提取论文中所有学者信息，从中选出 h-index排名最靠前的 2000位领域活跃学者，分析了学者的分布等情况，介绍了部分该领域国内外知名度较高的活跃学者。

我们所研究的期刊和会议包括：

ACM Transactions on Computer-Human Interaction（TOCHI）

International Journal of Human Computer Studies（IJHCS）

ACM Conference on Computer Supported Cooperative Work and Social Computing（CSCW）

ACM Conference on Human Factors in Computing Systems（CHI）

ACM International Conference on Ubiquitous Computing（UbiComp）

ACM Symposium on User Interface Software and Technology（UIST）

3.1学者情况概览

□ 全球学者概况

学者分布地图对于进行学者调查、分析各地区竞争力现况尤为重要，图 3-1为人机交互领域全球顶尖学者分布状况。地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从地区角度来看，美国的东部人才最为集中，西欧也有较多的人才分布，少数的学者分布于亚洲东部，大洋洲、非洲、南美洲的学者稀少。

AMiner



图 3-1 人机交互领域全球学者分布

TOP5 国家分布如图 3-2所示。从国家角度来看，美国在人机交互领域的人才数量遥遥领先，英国次之，中国学者数量排名第五。

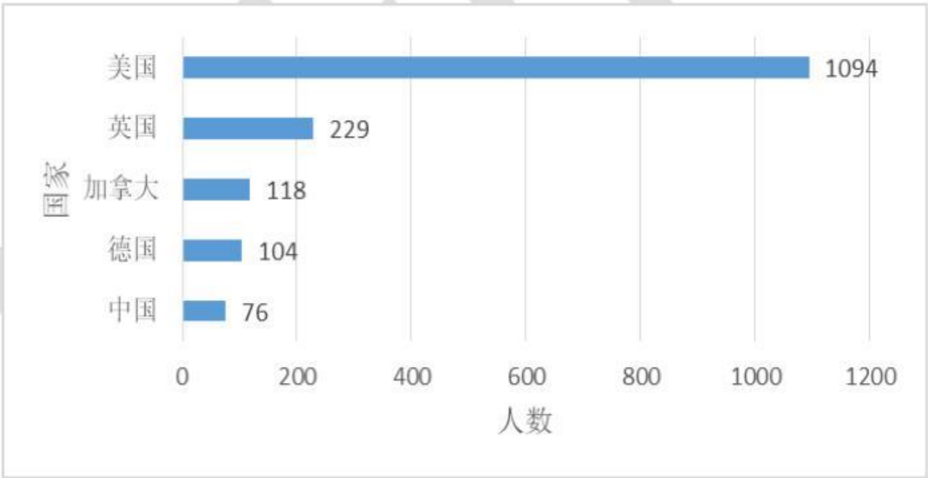


图 3-2 人机交互领域TOP5国家

此外，在性别比例方面，人机交互领域中男性学者占比 81.9%，女性学者占比 18.1%，男性学者占比远高于女性学者。

□ h-index分布

人机交互学者的 h-index分布如图 3-3所示，分布大体呈阶梯状，其中 h-index在在 10以下区间的人数最多，有 873人；50-60区间的人数最少，有 55人。

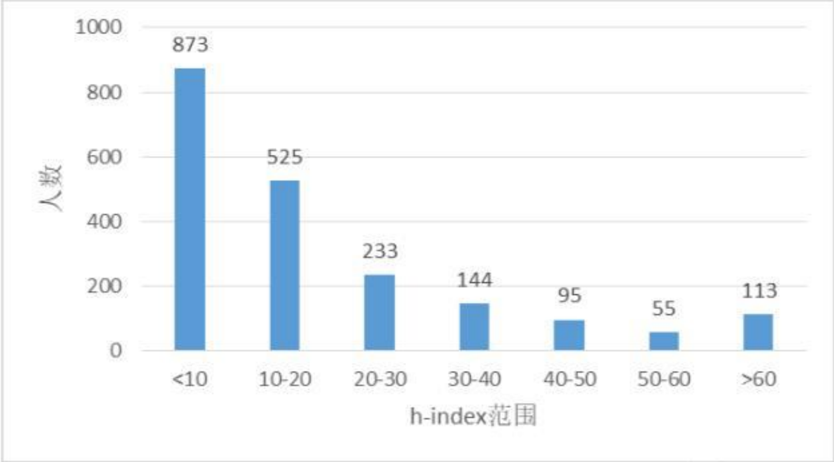


图 3-3 人机交互领域学者h-index分布

□ 中国学者分布

我国专家学者在人机交互领域的分布如图 3-4所示。可以看出，京津地区在本领域的人才数量最多，随后是台湾、长三角和珠三角地区，相比之下，中西部地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在人机交互领域学者数量方面优势不大。



图 3-4 人机交互领域中国学者分布

AMiner

□ 中外合作

中国与其他国家在人机交互领域的合作情况可以由 AMiner大数据平台分析得到，在上述期刊/会议中合作论文数量 TOP10的关系如图 3-5所示，从中可以看出中美合作的论文数遥遥领先，体现出中美在本领域的密切合作关系。

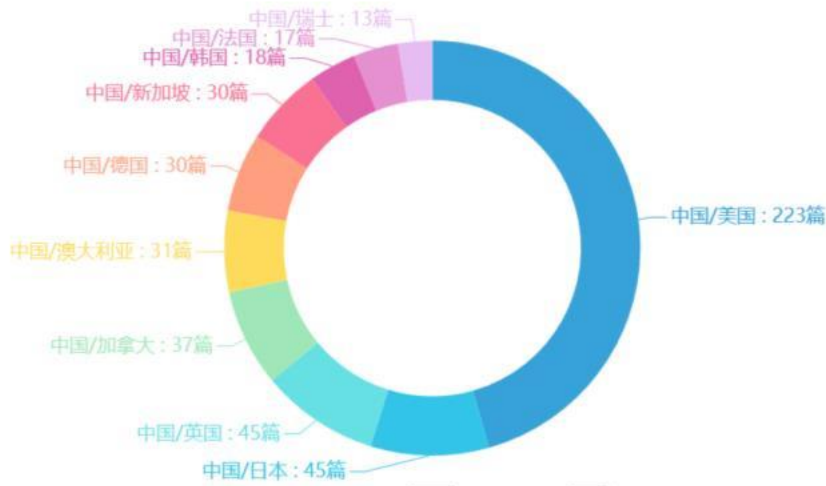


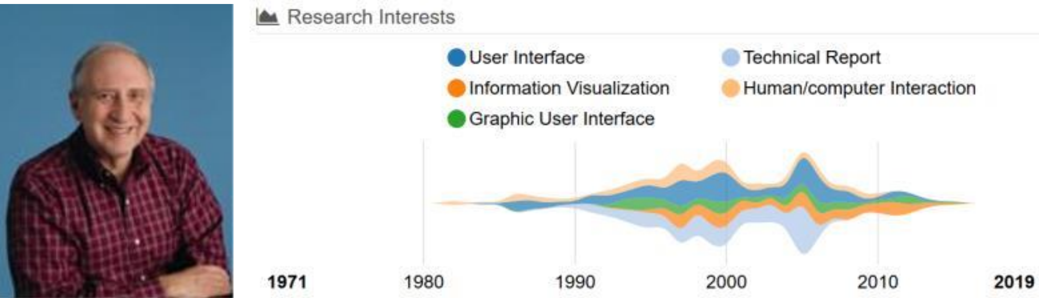
图 3-5 中国与其他国家的合作论文数量情况

3.2代表性学者简介

3.2.1国外代表性学者

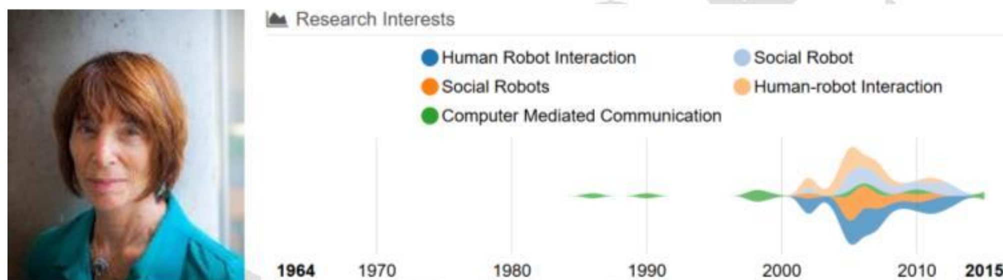
本节对人机交互领域部分国外高 h-index学者进行简要介绍，限于报告篇幅，我们不能对所有学者逐一罗列，如有疏漏，还请与 AMiner编者联系，或者登录 https://www.aminer.cn/获取更多资料。

□ **Ben Shneiderman**



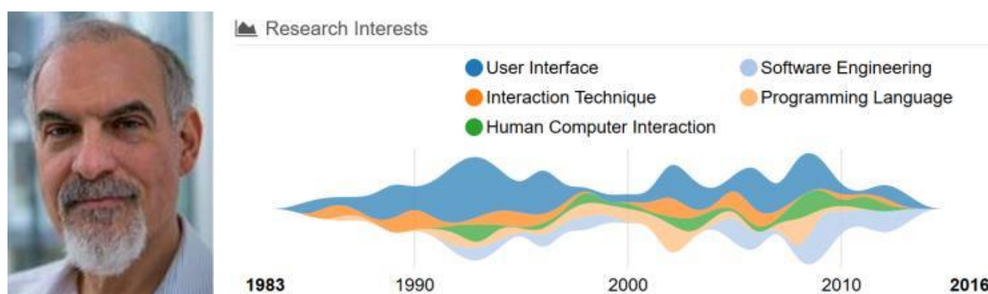
Ben Shneiderman（h-index：137），1973年在纽约州立大学石溪分校获得心理学博士学位，马里兰大学计算机科学系教授，马里兰大学人机交互实验室主任，ACM Fellow，AAAS Fellow，IEEE Fellow。他的研究兴趣是计算机科学，人机交互，信息可视化社交媒体等。Shneiderman 是美国国家工程院院士，美国国家发明家学会院士。他编写了 *Leonardo's Laptop: Human Needs and the New Computing Technologies*等多部著作，其中 *Designing the User Interface: Strategies for Effective Human-Computer Interaction*一书被翻译成 6种语言，列出了人机交互课程中经常教授的“界面设计的八条黄金法则”。曾获得 IEEE Visualization Career Award、SIGCHI Life Time Achievement、Miles Conrad Award等荣誉。

- **Sara Kiesler**



Sara Kiesler（h-index：93），1965 年在俄亥俄州立大学获得心理学博士学位，卡内基梅隆大学人机交互研究所教授，ACM Fellow，American Psychological Association Fellow，Psy Chi Fellow。她的研究兴趣广泛，从计算机介导的通信、计算机支持的协作工作到人机交互，曾获得 CHI Lifetime Achievement Award、International Communication Association Williams Prize、InGROUP McGrath Lifetime Award、Human Robot Interaction Lifetime Service Award等荣誉。

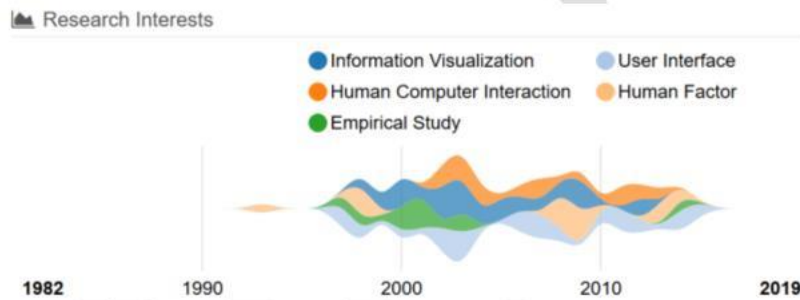
- **Brad A. Myers**



Brad A. Myers（h-index：87），1987年在多伦多大学获得了计算机科学博士学位，师从比尔·巴克斯顿，卡内基梅隆大学人机交互研究所的教授， ACM Fellow，IEEE Fellow。他的研究兴趣包括用户界面、编程环境、编程语言设计、最终用户软件工程、API可用性、开发人员体验、交互技术、示例编程、手持计算机和可视化编程等。Myers 曾获得 15 个最佳论文类型奖和 5 个最具影响力论文奖。2017 年，他因在人机交互研究方面的杰出贡献，获得了 ACM SIGCHI Lifetime Achievement Award。

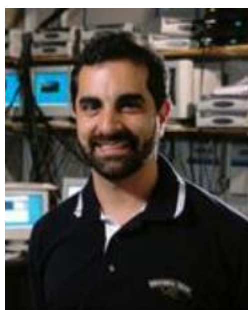
- **Mary Czerwinski**

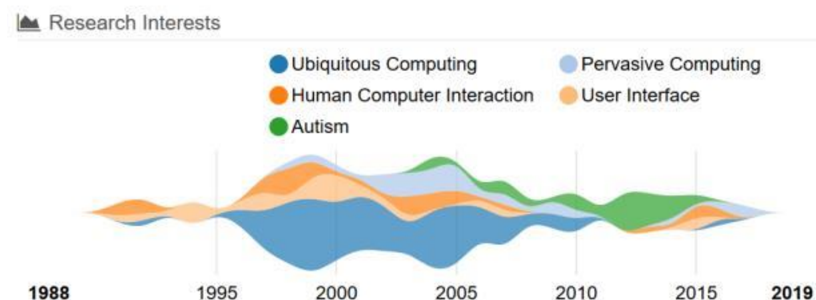




Mary Czerwinski（h-index：84），在印第安纳大学获得了认知科学博士学位，美国认知科学家和人机交互专家，微软研究院担任可视化和交互研究小组经理，ACM Fellow， American Psychological Science association Fellow，European Alliance for Innovation Fellow。她的研究主要集中在情绪跟踪、信息工作者任务管理、多任务处理以及个人和群体的意识系统，曾获得 ACM SIGCHI Lifetime Service Award、Distinguished Alumni award from the College of Arts and Sciences from Indiana等荣誉。

- **Gregory D. Abowd**





Gregory D. Abowd（h-index：82），1987年在牛津大学获得理学硕士学位和哲学博士学位，佐治亚理工学院交互计算学院教授，ACM Fellow。他主要从事人机交互、泛在计算、软件工程和计算机支持的协作工作等领域的研究。Abowd是亚特兰大当地一个非营利性组织——亚特兰大自闭症协会（Atlanta Autism Consortium）的创始主席，曾获得 ACM SIGCHI Social Impact Award、ACM Eugene Lawler Humanitarian Award等荣誉。

- **John M. Carroll**

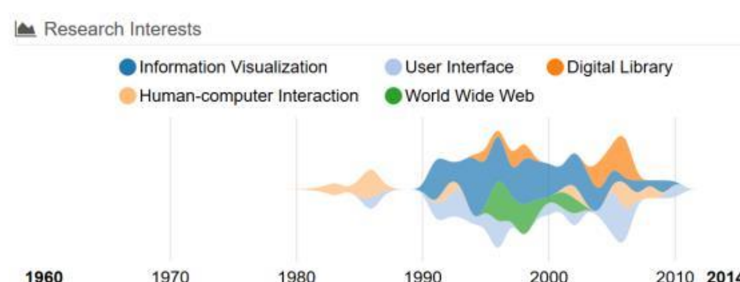




John M. Carroll（h-index：81），1976年在哥伦比亚大学获得实验心理学博士，宾西法尼亚州立大学信息科学与技术杰出教授，ACM Fellow，IEEE Fellow。他是人机交互研究的创始人之一，在过去的二十年里一直致力于人机交互领域的发展。Carroll曾在 1982年的计算系统人为因素标准局会议的项目委员会任职，这实际上是该领域旗舰系列会议——*ACM CHI Conferences*的前身。他是该领域主要期刊 *ACM Transactions on Computer-Human Interaction*的创始副主编，也是*Transactions on Information Systems*和 International Journal of Human-Computer Interaction编辑委员会的创始成员，曾获得 ACM Joseph T. Rigo Award、ACM SIGCHI Lifetime Achievement Award等荣誉。

- Stuart Card

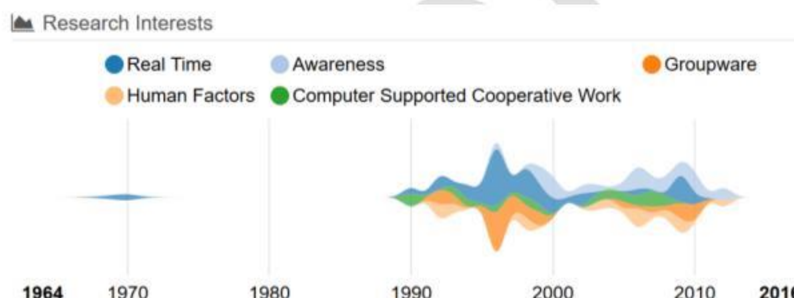




Stuart Card（h-index：80），在卡内基梅隆大学获得心理学博士学位，施乐公司高级研究员，ACM Fellow。他的研究兴趣是人机交互的理论描述，包括模型人处理器、用户交互的 GOMS理论、信息搜集理论和互联网使用的统计描述。Card 是 *The Psychology of Human-Computer Interaction*、*Human Performance Models for Computer-Aided Engineering*一书的合著者，并在许多编辑委员会任职。曾获得 CHI Lifetime Achievement Award、The Franklin Institute's Bower Award and Prize for Achievement in Science等荣誉。

- **Saul Greenberg**

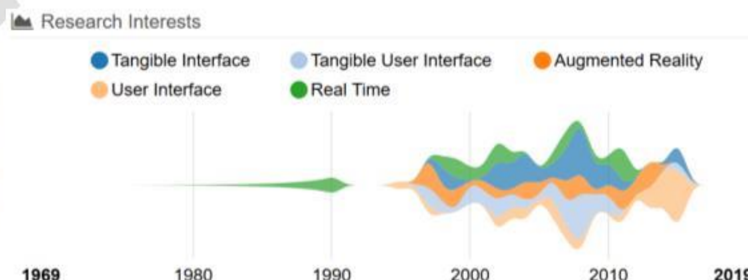




Saul Greenberg（h-index：78），1989年在卡尔加里大学获得博士学位，卡尔加里大学计算机科学系的教授，ACM Fellow。他的研究领域涉及人机交互、计算机支持的协作工作和普适计算等方面。曾获得 Canadian Human Computer Communications Society Achievement Award、the ACM UIST Lasting Impact Award、ACM ISS 10-Year Impact Award等荣誉。

- **Hiroshi Ishii**



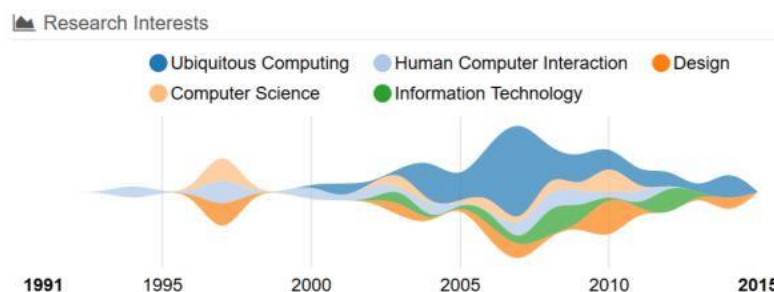


Hiroshi Ishii（h-index：76），在北海道大学获得电子工程学士学位、计算机工程硕士以及计算机工程博士学位，麻省理工学院教授和媒体实验室副主任。他

主要从事人机交互、计算机支持协同工作相关领域的研究。Ishii 领导了 NTT人机交互实验室的计算机支持协同工作小组，2001年获得麻省理工学院终身教职。他是人机交互有形用户界面领域的先驱者，公认的 TUI创始人。2019年 Ishii获得了 SIGCHI Lifetime Research Award，以表彰他过去 25年为人机交互领域作出的卓越贡献。

- **Paul Dourish**

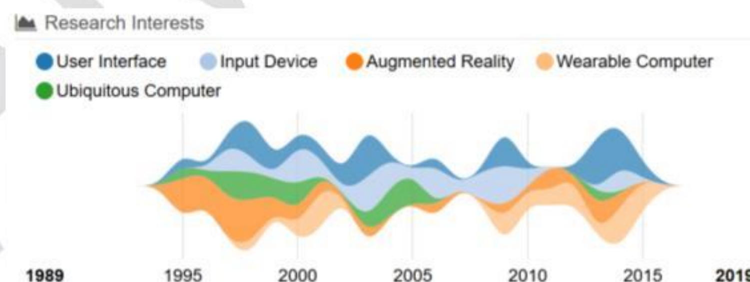




Paul Dourish（h-index：73），在伦敦大学学院的计算机科学博士学位，加州大学欧文分校信息学教授，ACM Fellow。他的研究兴趣包括泛在计算、计算机支持的协作工作和科学技术的社会研究。曾获得 Diana Forsythe Prize、BM Faculty Award、National Science Foundation CAREER Award等荣誉。

- **Jun Rekimoto**



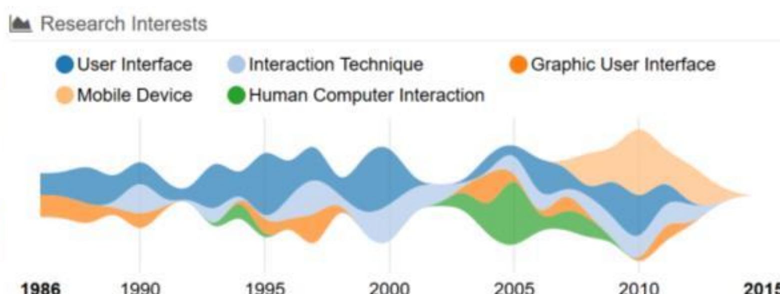


Jun Rekimoto（h-index：71），1996年在东京工业大学获得博士学位，东京大学信息学系教授，Koozyt 公司的共同创始人，同时是索尼计算机科学实验室副主任。他主要从事人机交互、增强现实（AR）、虚拟现实（VR）、人体增强、远程呈现以及混合智能系统相关领域的研究。Rekimoto 在移动设备和计算、虚拟现实和信息可视化等领域发表了数百篇论文。曾获得 Multi-Media Grand Prix

Technology Award、Japan Inter-Design Award、ACM UIST Lasting Impact Award 等荣誉。

- **Scott Hudson**

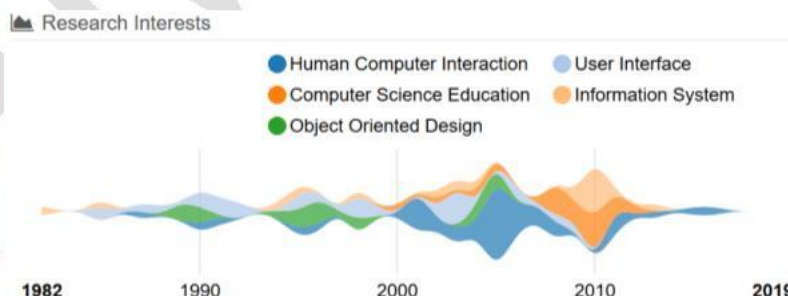




Scott Hudson（h-index：68），于 1986年在科罗拉多大学获得计算机科学博士学位，卡内基梅隆大学人机交互研究所教授，曾在亚利桑那大学任助理教授、佐治亚理工学院任计算机学院副教授。他主要从事交互技术、用户界面、可视化技术以及最终用户编程系统相关领域的研究。 Hudson 于 2006 年当选为 CHI Academy 成员，并定期在 *ACM SIGCHI* 和 *UIST* 会议项目委员会任职，同时是 *ACM Transactions on Computer-Human Interaction*的副编辑。他与同事共同开发了 Lumitrack（一种运动跟踪技术），已在视频游戏控制器和电影行业中使用。

- **Mary Beth Rosson**





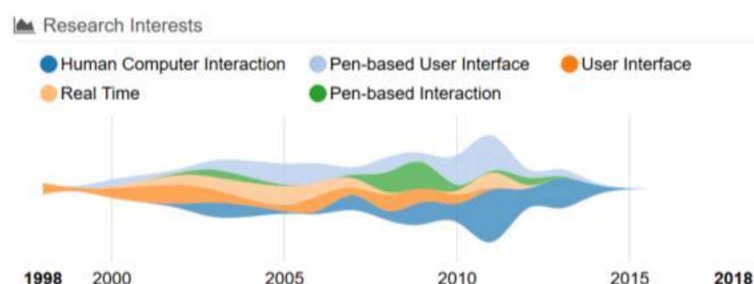
Mary Beth Rosson（h-index：67），于 1982年在德克萨斯大学奥斯汀分校获得实验心理学博士学位，宾夕法尼亚州立大学信息科学与技术学院教授。她主要从事最终用户编程、可视化编程、计算机支持协同工作以及人机交互相关领域的研究。Rosson 曾担任 *CHI 2007* 等著名会议的主席，目前是 *Human-Computer Interaction* 和 *Interacting with Computers* 的编辑,曾获得 ACM Recognition of Service Award、ACM Distinguished Scientist等荣誉。

3.2.2国内代表性学者

综合领域知名度与活跃度等方面，本节对国内人机交互领域部分代表性的学者进行简要介绍，排名不分先后，如有疏漏，还请与 AMiner编者联系，或者登录 https://www.aminer.cn/获取更多资料。

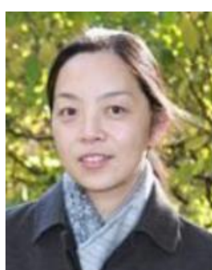
□ 戴国忠

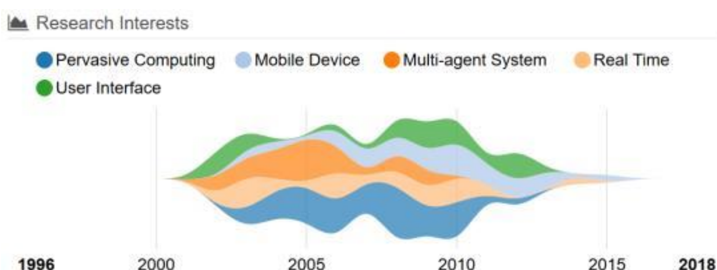




戴国忠，1962 到 1967 在中国科学技术大学应用数学系学习，1970 到 1982 在中国科学院自动化研究所从事相关研究工作，1982到 1985作为访问学者在美国马里兰大学计算机系研习，1985回到中国科学院自动化研究所，1990到 1995 年担任北京软件工程研制中心副主任，现担任中国自动化学会计算机图形学和 CAD专业委员会主任、国家八六三计划 CIMS主题专家组专家、浙江大学 CAD 和计算机图形学国家重点实验室学术委员会成员、中国科学院软件研究所总工程师、博士生导师。戴国忠主要从事计算机图形学、计算机用户界面和人机交互相关领域的研究。他是我国计算机用户界面领域的开拓者之一，同时是 ACM人机交互学会中国分会主席，中国自动化学会计算机图形学和 CAD专业委员会主任，曾获得国务院特殊津贴，全军科技进步二等奖，中国科学院科技进步一等奖、二等奖、三等奖，国家科技进步二等奖等荣誉。

□ 史元春

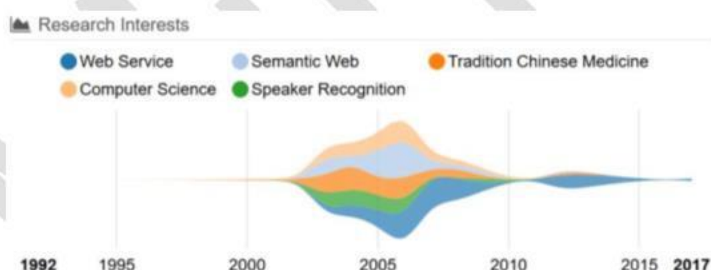




史元春，在清华大学计算机系获得学士、硕士、博士学位，清华大学计算机系“长江学者”特聘教授，博士生导师，清华大学全球创新学院 GIX 院长，清华大学计算机系人机交互与媒体集成研究所所长，清华信息科学与技术国家实验室普适计算研究部主任，清华大学计算机系学术委员会副主任，中国计算机学会常务理事，普适计算专业委员会主任，中国计算机学会常务理事，中国图象图形学学会多媒体专业委员会委员，IEEE 高级会员，主持了国家十三五重点专项等重大国家科技计划项目。史元春的研究方向主要为人机交互、普适计算、网络多媒体、eLearning 等相关领域，其科研成果包括交互效率模型、自然动作建模方法、智能空间计算平台、多媒体多播协议等。她曾任 *CCCF*副主编，*Interacting with Computers*等国际期刊编委。近年连续获得国际顶级会议最佳论文奖，两次获得国家科技进步奖。

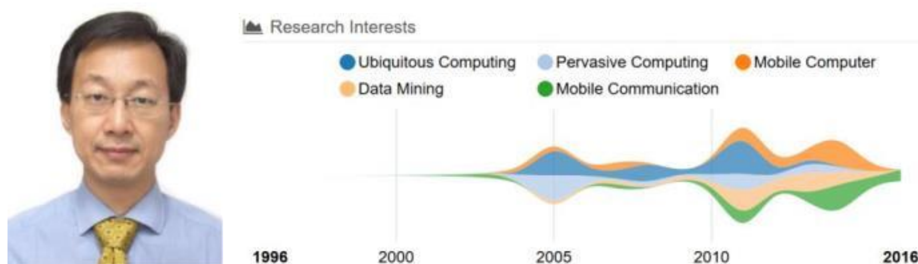
□ 吴朝晖





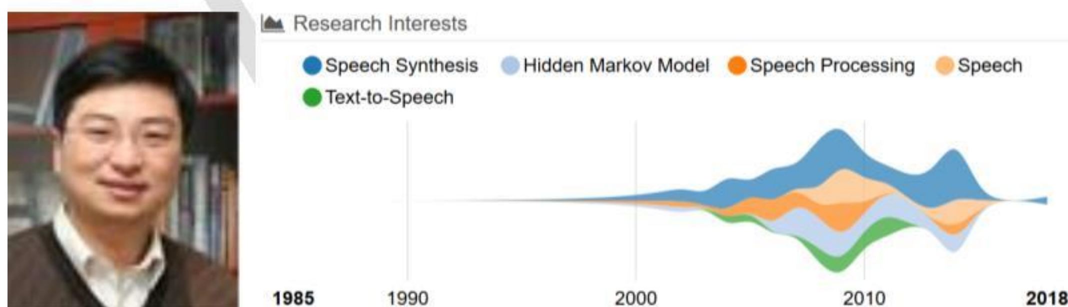
吴朝晖，1993年在浙江大学获博士学位，浙江大学校长、教授、博导，2017年当选为中国科学院院士。主要从事计算机领域研究，研究兴趣包括人工智能、服务计算、普适计算、嵌入式系统、混合智能、脑机接口等。在服务计算方面，他提出了复杂服务计算理论模型与方法，主持研制了复杂服务计算支撑平台，产生了重大经济与社会效益。在脑机融合的混合智能方面，吴朝晖提出了混合智能体系结构及脑在回路的信息处理模型，发明了多种脑机间智能交互、融合增强的技术方法，主持构建了听视觉增强的大鼠机器人等原型系统，产生了重要国际影响。作为第一完成人，吴朝晖曾获国家技术发明二等奖 1项、国家科技进步二等奖 1项、何梁何利科技创新奖，研究成果入选 2016年中国高校十大科技进展。

□ 张大庆



张大庆，1996 年获得意大利罗马大学博士学位，北京大学讲席教授，中国计算机学会普适计算专委会副主任。主要研究领域包括普适计算、移动计算、城市计算等。张大庆目前担任 *IEEE Pervasive Computing*，*IEEE Trans. on Big Data*，*ACM IMWUT*，*ACM Trans. on Intelligent Systems and Technology*等多个国际顶级期刊的编委，也曾担任过 10 余个知名国际会议的大会或程序委员会主席，应邀在 20多个国际会议做大会特邀报告。他带领的北大团队新提出的基于菲涅尔区无线感知理论获得了国际同行的广泛认可，在利用商用 WIFI支撑呼吸检测、室内跟踪等一系列应用中均做到国际性能最佳。其所创的情境感知模型被国际普适计算、移动计算和服务计算学术界所广泛采用，并被普适计算领域顶级会议 *IEEE PerCom 2013*授予“十年最具影响力论文奖”。近年来，张大庆在群智感知、感知大数据分析和无接触感知等新兴研究方向工作，先后获得中国计算机学会推荐的全部 4个普适计算国际会议的最佳论文或提名奖。

□ 陶建华



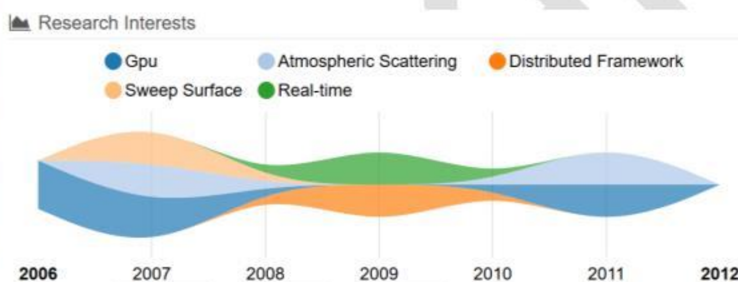
陶建华，拥有南京大学电子系学士学位、南京大学声学专业硕士学位、清华大学计算机专业博士学位，中国科学院自动化研究所研究员，模式识别国家重点实验室副主任，博士生导师，国家杰出青年基金获得者，国家“万人计划”科技

AMiner

创新领军人才，中国计算机学会会士和常务理事，中国人工智能学会理事，中国中文信息学会理事，中国图象图形学学会人机交互专委会主任。陶建华的研究领域主要为语音合成与识别、多媒体信息处理、人机交互、模式识别。先后负责和参与国家级项目 40 余项。他是 *IEEE Trans. on Affective Computing* 编委、*International Journal on Multimodal Interfaces* 等刊物的副编辑。陶建华在包括 *IEEE Trans.*、*ICASSP*、*ICSLP*、*ICCV*、*ICIP*等国内外学术期刊和会议上发表论文 200余篇，论文和成果曾获全国人机语音通讯学术会议、全国信号处理会议、全国人机交互会议优秀论文、北京市科技进步二等奖等。

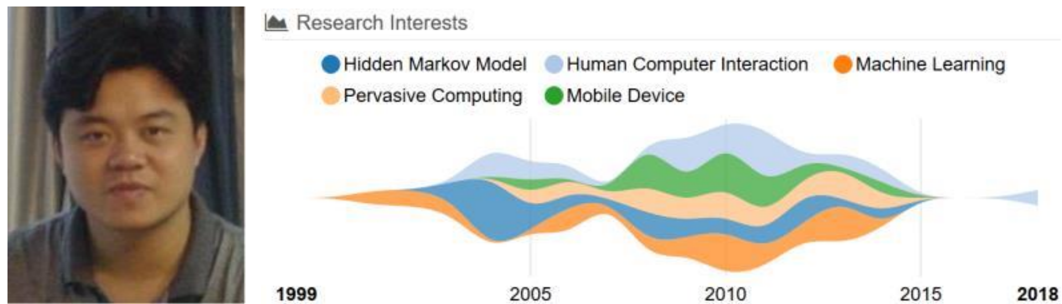
□ 汪国平





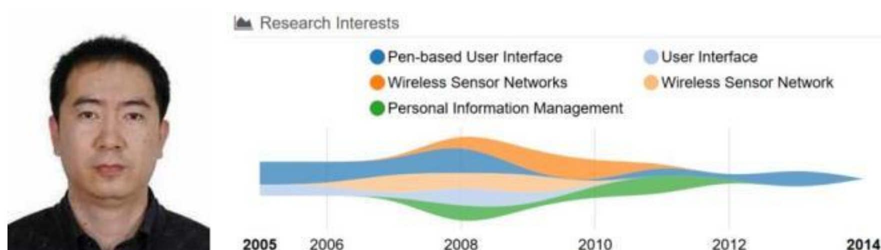
汪国平，1997 年获得复旦大学理学博士学位，现任北京大学图形与交互技术实验室主任，北京大学信息学院软件研究所副所长，兼任中国计算机学会虚拟现实和可视化专业委会副主任委员、中国图象图形学会常务理事、北京图象图形学会副理事长、国家信息标准化委员会委员、《计算机辅助设计与图形学学报》编委等。汪国平主要从事计算机图形学、虚拟现实与仿真、人机交互等方面的教学与研究。他近几年来负责 973课题、863项目、国家攻关计划项目和国家自然科学基金等各种国家级和省部级项目三十余项。2001 年其负责研制了国内第一个基于 AICC和 SCORM国际标准的在线多媒体学习系统，已经在佳能公司和国内外多个大中型企业获得良好应用。申请国家发明专利 13项，授权 8项，在本专业顶级期刊 *CAD*、*CAGD*、*IEEE*和 *ACM*等国内外重要学术期刊和学术会议上发表学术论文 120余篇，其中有两篇论文分别在 *ACM SPM2007*和第十届 *IEEE CAD/Graphics*国际会议上获得最佳论文奖。2008年获得教育部科技进步一等奖，2009年获得国家杰出青年基金，2010年获国务院政府特殊津贴。

□ 陈益强



陈益强，拥有湘潭大学计算机应用专业学士学位、计算机软件与理论专业硕士学位以及中国科学院计算技术研究所计算机应用专业博士学位，中国科学院计算技术研究所研究员，先后担任了中科院计算所上海分所常务副所长、中科院计算所普适计算研究中心副主任、中科院计算所普适计算研究中心主任。他主要从事人机交互与普适计算相关领域的研究。陈益强是 IEEE会员、CCF高级会员、YOCSEF AC 委员、IEEE 多媒体系统及应用技术委员会委员，是中国计算机学会普适计算专业委员会委员、中国图形图象学会多媒体专业委员会委员、中国标准化协会多媒体广播技术委员会委员，也是 *IEEE TCSVT*、*IEEE TKDE*等多种国际期刊审稿人。陈益强教授主持了国家自然科学基金重大研究计划子课题、国家科技支撑计划子课题、国家高科技 863计划项目以及国家自然科学基金与工信部电子发展基金等多项课题的研究，在国际国内期刊和会议上发表论文八十余篇，申请多项专利。他曾荣获 2002年度中国科学院院长奖优秀奖、2003年度国家科技进步奖二等奖、2004 年度首届中国科学院优秀博士论文奖、2005 年度全国百篇优秀博士论文提名奖、2015年度北京市科学技术奖、2017年度中国计算机学会科学技术奖等。

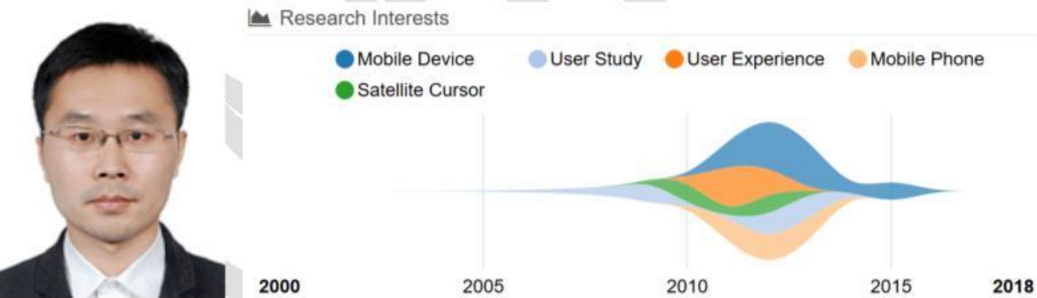
□ 田丰



AMiner

田丰，拥有西北大学计算机学士学位、硕士学位以及中国科学院软件研究所计算机应用技术博士学位，中国科学院软件研究所研究员、所长助理，国家重点研发计划首席科学家，国家优秀青年基金获得者，现担任 ACM SIGCHI中国分会主席、中国计算机学会人机交互专委会秘书长、中国人工智能学会智能交互专委会副主任、中国图象图形学学会人机交互专委会副主任等职。田丰长期从事人机交互相关研究工作，在笔式界面模型和方法、复杂笔迹结构理解和智能交互技术，笔式人机交互关键应用等方面取得了代表性研究成果。他是 *ACM Transactions on Interactive Intelligent Systems*、*Journal of Visual Languages and Computing*等人机交互领域著名国际刊物编委，在 *CHI*、*IJHCS*，*TPAMI*、*CSCW* 等国内外重要期刊会议上录用/发表学术论文百余篇，合作出版专著 2部、译著 1部。田丰作为第一完成人获得了 2015年度北京市科学技术一等奖、2018年度国家科技进步二等奖。

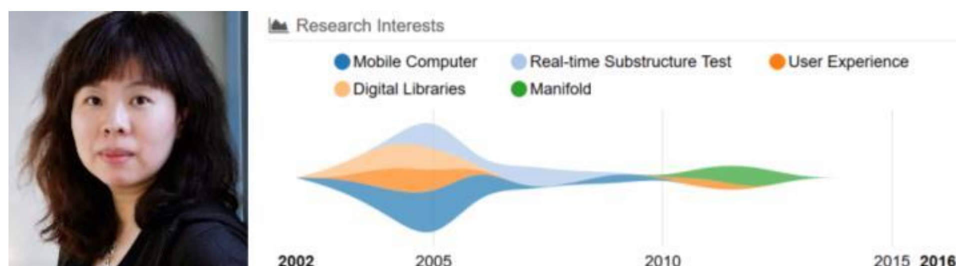
□ 喻纯



喻纯，2012 年在清华大学获得博士学位，清华大学计算机科学与技术系副教授。他的研究方向是人机交互，其研究围绕人机交互自然性的计算原理和优化方法开展，具有重要的理论和实际应用价值。喻纯创新性地将信息论和概率统计的方法应用在交互效率和自然性的建模优化中，其具体研究成果包括基于信息熵的任务复杂度度量、人体运动控制参数的统计建模、自然交互意图的贝叶斯推理等。他的研究获得自然科学基金青年基金，自然科学基金面上项目的支持，并参与了十三五重点研发计划前沿基础类项目“人机交互自然性的计算原理”和共性关键技术类项目“云端融合的自然交互设备和工具”，在两个项目中均为核心骨干，承担关键科研任务。此外，其研究也受到英特尔、三星、微软、搜狗、华为等企业的支持和认可。目前，多项研究成果正在转化过程中。喻纯在人机交互领

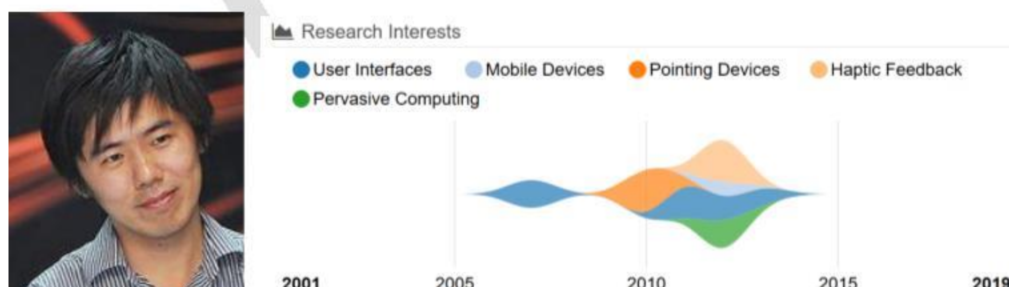
域高水平期刊和会议上发表论文 30余篇，包括 12篇 CCF A类论文，近年来共获得 6次论文奖励，包括 *CHI 2013*、*MobileHCI 2013*、*CHI 2015*、*CHI 2016*、*CHI 2017*的最佳论文提名奖和 *IEEE UIC 2014*的最佳论文奖。他在 2015年获得教育部科技进步一等奖，2016年获得了 CCF-Intel青年学者计划的奖励资助。

□ 王茜莺



王茜莺，拥有上海交通大学电子工程专业学士学位、电子工程和工业管理双硕士学位、美国斯坦福大学人机交互专业博士学位，现任联想集团副总裁，联想研究院技术战略与创新平台以及智慧教育产品研发平台总经理、ACM 人机交互分会中国副主席。王茜莺带领团队一向以创新能力突出著称，连续多年获得联想最佳创新团队称号，共获得超过 80 篇发明专利并在知名学术期刊上发表过 20多篇论文。她在 IBM中国研究中心工作期间，参与了 IBM ThinkScribe/Crosspad DBCS以及 ThinkPad TransNote DBCS等产品的设计与研发，后者获得了由商业周刊颁发的 2001 年度美国优秀工业设计奖（IDEA），两度获得 IBM 全球研发系统成就奖，2019年荣登福布斯“中国科技女性榜”。

□ 曹翔



曹翔，先后在多伦多大学攻读人机交互硕士、博士学位。小小牛科技创始人及 CEO，曾作为微软的研究员，先后在英国剑桥的实验室（MSRC）、北京实验室（MSRA）从事相关研究，他在微软工作了 5 年，期间参与设计过 Tell Table

故事桌、Kinect体感游戏等产品。曹翔还曾在西门子，IBM和英特尔工作和实习，并在剑桥大学的科珀斯克里斯蒂学院任研究员。他的研究工作涉及人机交互的多个领域，最著名的是新颖的交互技术和设备以及支持创造力的数字工具，以及计算机支持的合作活动和用户绩效建模。曹翔是 *Springer*人机交互系列书籍编辑委员会委员，众多会议和期刊的论文审稿人，例如 *CHI*、*UIST*、*CSCW*、*TOCHI*、*SIGGRAPH*，还是 IEEE国际信息和自动化会议（ICIA）的项目共同主席。他是 *ACM Special Interest* 中国分会的副主席，被国际中国计算机人机交互协会评为HCI杰出青年领袖。

3.2.3近三年高产学者

本节对近三年在本报告所选期刊/会议上发表论文数最多的 100 名学者进行统计，如表 3-1所示（截止到 2020年 2月）。

表 3-1 近三年高产学者百人名单

序号	姓名	任职机构	论文数
1	Chun Yu	清华大学	30
2	Yuanchun Shi	清华大学	27
3	Meredith Ringel Morris	微软研究院	25
4	Kasper Hornbæk	哥本哈根大学	23
5	John Vines	诺森比亚大学	23
6	Chris Harrison	卡耐基梅隆大学	23
7	Vassilis Kostakos	墨尔本大学	20
8	Christian Holz	苏黎世联邦理工学院	20
9	Xing-Dong Yang	达特茅斯学院	20
10	Daniel Vogel	滑铁卢大学	19
11	Carl Gutwin	萨斯喀彻温大学	19
12	Niels Henze	斯图加特大学	19
13	Florian Alt	慕尼黑联邦国防军大学	19
14	Tovi Grossman	多伦多大学	19
15	Patrick Olivier	莫纳什大学	17
16	Carman Neustaedter	西蒙弗雷泽大学	17
17	Daniel Wigdor	多伦多大学	17
18	Patrick Baudisch	哈索普拉特纳研究所	17
19	Ken Hinckley	微软研究院	16
20	Steve Benford	诺丁汉大学	16
21	Sriram Subramanian	萨塞克斯大学	16

序号	姓名	任职机构	论文数
22	Michael S. Bernstein	斯坦福大学	16
23	Wendy E. Mackay	法国国家信息与自动化研究所	16
24	Neha Kumar	佐治亚理工学院	16
25	Karrie Karahalios	伊利诺伊大学香槟分校	16
26	Jürgen Steimle	萨尔大学	16
27	Shaun W. Lawson	诺桑比亚大学	15
28	Jorge Gonçalves	墨尔本大学	15
29	Munmun De Choudhury	佐治亚理工学院	15
30	Albrecht Schmidt	慕尼黑大学	15
31	Rob Comber	瑞典皇家理工学院	15
32	Jennifer Mankoff	卡耐基梅隆大学	15
33	George W. Fitzmaurice	欧特克研究中心	14
34	Peter C. Wright	纽卡斯特大学	14
35	Xiaojun Bi	纽约州立大学石溪分校	14
36	Sean Follmer	斯坦福大学	14
37	James A. Landay	斯坦福大学	14
38	Scott E. Hudson	佐治亚理工学院	14
39	Jun Gong	达特茅斯学院	14
40	Yong Li	清华大学	14
41	Uichin Lee	韩国高等科学技术学院	13
42	Volker Wulf	锡根大学	13
43	Anne Roudaut	布里斯托大学	13
44	Antti Oulasvirta	阿尔托大学	13
45	Pawel W. Wozniak	乌得勒支大学	13
46	Jacob O. Wobbrock	华盛顿大学	13
47	Yunan Chen	加州大学欧文分校	13
48	Yang Zhang	卡耐基梅隆大学	13
49	Bing-Yu Chen	国立台湾大学	13
50	Gregory D. Abowd	佐治亚理工学院	13
51	Andy Cockburn	坎特伯雷大学	12
52	Géry Casiez	里尔大学	12
53	Frank Vetere	墨尔本大学	12
54	Daqing Zhang	北京大学	12
55	Xin Yi	清华大学	12
56	Niels van Berkel	奥尔堡大学	12
57	Brent J. Hecht	西北大学	12
58	Michel Beaudouin-Lafon	巴黎第十一大学	12
59	Lining Yao	卡耐基梅隆大学	12
60	Antonio Krüger	德国人工智能研究中心	12
61	Stephen A. Brewster	格拉斯哥大学	12

AMiner

序号	姓名	任职机构	论文数
62	Katharina Reinecke	华盛顿大学	12
63	Madeline Balaam	瑞典皇家理工学院	12
64	Maneesh Agrawala	斯坦福大学	12
65	Wendy Ju	康奈尔大学	12
66	Jeffrey P. Bigham	卡耐基梅隆大学	12
67	Mor Naaman	康乃尔大学	12
68	Bongshin Lee	微软研究院	12
69	Nathalie Henry Riche	微软研究院	12
70	Philip J. Guo	加州大学圣地亚哥分校	12
71	Gierad Laput	苹果公司	12
72	John M. Carroll	宾夕法尼亚州立大学	12
73	James Fogarty	华盛顿大学	12
74	Pamela J. Wisniewski	中佛罗里达大学	11
75	Matt Jones	斯旺西大学	11
76	Sean A. Munson	华盛顿大学	11
77	Giulio Jacucci	赫尔辛基大学	11
78	Shamsi T. Iqbal	微软研究院	11
79	Tawanna R. Dillahunt	密歇根大学	11
80	Eyal Ofek	微软研究院	11
81	Julie A. Kientz	华盛顿大学	11
82	Alexis Hiniker	华盛顿大学	11
83	Pourang Irani	曼尼托巴大学	11
84	Kellie Morrissey	纽卡斯尔大学	11
85	Andreas Bulling	斯图加特大学	11
86	Enrico Rukzio	乌尔姆大学	11
87	Mayank Goel	卡耐基梅隆大学	11
88	Da-Yuan Huang	国立交通大学	11
89	Liwei Chan	国立交通大学	11
90	Amanda Lazar	马里兰大学	10
91	Desheng Zhang	罗格斯大学	10
92	Eric Gilbert	密歇根大学	10
93	Florian Floyd Mueller	墨尔本皇家理工学院	10
94	François Guimbretière	康奈尔大学	10
95	Geraldine Fitzpatrick	维也纳工业大学	10
96	Jaime Teevan	微软公司	10
97	Jason C. Yip	华盛顿大学	10
98	Jason I. Hong	卡耐基梅隆大学	10
99	Jeffrey Bardzell	印第安纳大学	10
100	Jennifer Pearson	斯旺西大学	10

4应用篇

4.1智能终端

□ 智能手机

手机人机交互一直都是影响用机体验的重要因素，手机的交互方式也在不断进化。从最初的实体按键，到屏幕触控，人机交互进行了全面升级。虽然屏幕触控仍旧是目前主流且重要的交互方式，但手机厂商仍旧在不断进行创新和尝试，给用户带来更优秀的使用体验。

在功能机时代，用户与手机的交互基本都是通过实体按键。智能手机时代的到来，让人机交互思路有了非常大的改变，朝着触屏控制的方向发展，手机上剩下的实体按键变得越来越少。虚拟导航键的出现取代了实体导航键，虚拟导航键直接以画面的形式显示在屏幕上，可以实现同样的功能。这种导航键设计不会占用机身的空间，使用起来也很方便，但会占用屏幕显示空间。在手机上加入指纹识别、人脸识别等可以简化操作步骤，原本需要输入密码才能解锁，现在只需指纹按压识别就能直接实现解锁，这些变化都让日常使用变得更加方便。为了提升交互效率，许多手机还内置了智能语音助手，通过简单的语音指令便能代替以往复杂的操作，甚至还有很多厂商给手机加上了独立的 AI语音按键，可以一键直达语音助手。虽然目前对于很多人来说智能语音交互只是个辅助方式，但这也是一种非常重要的交互方式。相信未来语音交互将变得更好用，通过语音助手所能实现的功能也会更多[62]。

□ 车载智能终端

汽车已从单纯的代步工具，逐步发展为集出行、办公、生活、娱乐于一体的生活伴侣，汽车人机交互也从原来以实体操作装置为基础，注重功能实现的机械交互方式，发展为以关注人的行为为基础，包括触控、语音、手势、眼动等多通道交互技术在内的自然交互方式。

目前触控交互主要以触屏交互为主，即驾驶员在中控屏幕上通过手势触控操作。与机械交互不同，触屏交互没有明显的物理反馈，无形中增加了驾驶员感知负荷。为解决这样的问题，奥迪在 A8车型上应用了触觉反馈技术，即驾驶员需要施加 2.5牛的力才可以触发操作，同时有一个力反馈。目前语音识别应用较为广泛，是一种完全解放双手和视觉的交互方式，现阶段只是以简单命令式语音为主，还需要语音识别开发厂商通过对人工智能技术和对语义库的不断完善，来使语音识别逐渐向自然语言交互延展。手势识别也是自然交互的一个重要研究方向，它能够最大限度地减少认知与视觉交互成本，目前宝马已在实车上应用了此项技术，驾驶员可以通过手左右挥动、手指旋转等动作进行简单操控。拜腾也在CES展上推出了三维手势控制技术，车辆能够识别拖拽、OK等更为复杂的手势动作[63]。

□ 智能单兵终端

现代战争正由机械化向信息化转变，数字化战场逐渐形成。单兵系统作为其中的关键因素，是把单兵作为整个作战系统的一个武器平台，从人-机-环境整体考虑、统筹规划与设计的一个完善集成的人机系统[64]。士兵作为单兵系统的载体，其与计算机的智能结合是通过人机交互作用实现的，人机交互技术在该领域发挥着越来越重要的作用。

在资源信息理解和整合方面，单兵终端对各种输入设备提供的不同格式、不同意义的数据进行处理，通过采用图像处理技术对士兵的位置、手势等进行跟踪和识别，采用语音识别、自然语言理解技术对士兵的命令进行解释等。例如在英国“未来一体化士兵技术”（FIST）计划中的“电子个人武器”（EIW）系统，这种电子控制模块化武器带有语音传感器和液晶显示器，传感器用于监视武器装弹和枪管温度，液晶显示器用于提示过热和低电警告，并提醒步兵武器里面还剩下多少弹药，可调射速。士兵对着语音传感器说的话可进入计算机菜单，菜单中有打开保险、开火等程序，代替了以往士兵扣动扳机的操作，可避免误伤。

在任务处理的智能化方面，单兵终端可以进行更复杂、更智能的处理工作，通过多维信息的高效分析计算，为士兵提供智能化决策辅助，提高作战任务处理

能力。例如美国为“陆地勇士”开发的“目标单兵武器系统”（OICW）通过其火控系统实现了对目标的智能化处理。在使用榴弹发射器射击时，士兵首先利用激光测距仪/数字罗盘（LRF/DC）并结合全球定位系统，了解到观测物的方位、高度和距离，瞄准目标，然后采用弹道计算器确定弹道方案，并通过扳机护圈上的操作键选择战斗模式并对目标参数进行修正。

4.2智能穿戴

人机交互技术是可穿戴移动终端系统及相关应用的核心技术，也是近年来国内外的研究热点。可穿戴移动终端在军事、工业、医疗、航天航空等领域有着广泛的应用前景，近年来涌现了大量可穿戴移动终端的人机交互理论、技术、应用研究，为推进计算系统小型化、普适化、智能化起到了巨大的作用。

2001 年法国学者 Robles-De-La-Torre 等在 Nature 上发表的论文，为可穿戴移动终端实现裸指触觉再现提供了重要的理论依据。近年来，美国、芬兰、日本等国家都验证了多媒体终端裸指触觉再现的可行性。目前，触觉再现渲染方法研究主要集中在机械力触觉再现设备上，触觉交互界面的研究刚刚起步。

头盔显示器从 1966 年问世以来，首先在军事上发挥着重要作用，如美国 Honeywell公司 IHADSS头盔显示器。近几年，面向消费市场的民用头盔显示技术逐步发展起来，出现了很多商业化的产品，如 2013年 Oculus VR公司推出的 Oculus Rift，具有高分辨率、大视场角、轻质量的特点。在透视显示方面，微软公司在 2015 年发布了 Hololens，具有良好的显示性能，可以准确地实时跟踪用户的头部运动，提供即时的交互体验。

近年来使用机器学习进行手势识别和分类成为了主流解决方案，如 2014年微软研究院的 Krupka 等提出了判别式分类器，在深度图像和红外图像上实现了手势的快速准确分类。但是目前的算法大多仅面向室内的桌面型应用，同时训练速度较慢，尚未出现针对可穿戴增强现实应用的实时手势识别方法。

在自然语言识别方面，目前国际上也启动了多个研究项目，如美国 IBM 公司长期经营的“Watson”计算机系统项目，日本的“Todai”机器人项目，DARPA 的“Big Mechanism”研究项目。

近年来，生理计算已逐渐引起学术界和产业界的共同关注。*ACM SIGCHI*年会在 2010和 2011年，分别举办了关于生理计算的研讨会。国际上许多著名的大学和研究机构，目前也逐步在开展生理计算方面的研究。生理信号的采集已逐渐从传统传感器采集向纳米织物采集过渡，即通过设计新型纳米材料，在长期舒适的穿戴中智能化地采集心电、肌电等生理信号[65]。

4.3智能家居

目前各类智能家居产品在生活中越来越多，而人们对智能家居人机交互便捷性、高效性的要求也越来越高，人机交互成为了科学研究的重中之重。

- 手机交互

手机交互是目前最常见的智能家居交互方式。手机 APP 相比传统的手动交互，实现了远程控制和定时开关，如苹果、海尔、嘟嘟 E家等品牌的手机 APP，不仅拥有这两种功能，还能设置情景模式，实现一键多控，非常便捷，此外，通过手机摄像功能还可以远程监控室内环境。

- 跨屏幕交互

手机交互可能会存在一个问题，当你没带手机的时候，或者手机在包里不便拿出的时候，可以通过跨屏幕交互解决这个问题。跨屏幕控制，即在平板电脑、室内机、电视、电脑、手表等设备进行控制，这些设备未必都能成为智能终端，尤其是手表，但它们至少能让用户及时了解各类情况，让用户不会错过关键信息。

- 语音交互

目前智能家居的语音交互主要有两种形式：直接语音与间接语音。直接语音，即直接对设备说话，发出语音指令，代表产品有国外的亚马逊的 echo、谷歌的

AMiner



Google Home，国内的叮咚智能音箱与嘟嘟智能语音管家；间接语音，即通过一款中间设备和其他设备展开交互，代表产品为各类智能手机的语音助手。

□ 手势交互

手势交互，即智能家居设备通过感应用户体感，或通过摄像头识别手势来接收和实现用户发出的指令，比如挥手关灯。这种交互方式有一定的局限性，不能广泛适用于整个家庭场景，但在某些方面，它甚至比语音交互更方便[66]。

4.4游戏领域

电子游戏经过几十年的发展，最明显的趋势就是向“自然化交互方式”的转变。这里重点讲述“体感游戏”，它不同于传统按键控制，是一种用肢体动作来直接完成游戏的“自然人机交互”模式。基于体感技术的交互界面在游戏应用中有如下的优势：

□ 加强人类感觉和动作通道的能力，减轻转化的负担

过去，手柄、键盘和鼠标等输入设备极大地减缓了操控的快感，当你彻底丢弃它们，就能享受放手一搏的新体验。体感技术使人操作计算机的方式更接近于日常自然交流，大大降低了学习成本。例如，记住字母按键在键盘中的位置，在图形交互界面中记住“按钮”和普通图片的区别，什么样的文字是超链接，以及右键点击鼠标会出现选项菜单等都需要一定的认知转化过程。而体感技术采用人们日常交流中经常用到的方式，如手势、说话等，在游戏中，体感技术避免了将游戏动作转化为各种按键而产生的困惑，使人与计算机的交流更加流畅。

□ 提供更多操作的可能，提高游戏的趣味性和玩家的参与性

自然交互界面是一个多学科交叉的研究领域。它包括多点触摸、体感手势识别、语音识别、重力加速感应、脑电波识别等多种技术形式。这些新技术丰富了游戏的输入方式，给玩家更多发挥的空间。随着电脑移动化程度的提高，平板电脑、手机成为游戏的新战场，多点触摸代替鼠标点选、语音输入，重力加速感应

以及震动等反馈机制使游戏的操作更加自然直接。在 Kinect 推出之后，通过摄像头直接拍摄玩家的动作，大大增强了玩家的参与感。

 高质量的反馈机制给用户提供了更加真实、多维度的感官体验

佩戴式 3D眼镜的浸入式场景为玩家提供了身临其境的感觉，游戏体验更加真实；大尺寸屏幕或投影、立体声取代了曾经的 21寸电视机，产生更强烈的震撼效果；最重要的是高精度的识别反馈机制，例如强大的运算能力能够快速准确地识别语音并给予及时恰当的反馈，手势操作时即使不能真实地拿到某个物体，屏幕中的虚拟物体也会牢牢地被掌握，体感识别能使人在屏幕中实时看到自己的动作与游戏角色间的密切联系等。

 既要做到轻松娱乐，又要考虑玩家的身心健康

一般游戏方式对操作有依赖，还会对我们的生理健康产生一定的影响，如鼠标手、键盘手、肥胖、高血压等症状屡见不鲜。体感游戏的出现让运动和娱乐实现完美的结合[67]。

4.5教育领域

以信息技术高速发展为背景的现代教育，学生的学习兴趣得到了提高，主体性得到了较大的发挥。但是，不可否认，绝大部分多媒体信息系统具有明显的局限性，例如学生只能按照时间的流程，按电脑中程序设计的流向有限制地进行浏览和学习，这大大降低了学生学习的参与性和主动学习的热情。而虚拟现实技术能够创建与现实社会类似的环境，更容易地满足情景化及自然交互性的要求，从而弥补了多媒体技术的不足。

在信息传播技术迅猛发展的今天，现有的教育内容、教育手段和教育方法正受到极大的挑战。虚拟现实技术能形象生动地表现教学内容，有效地营造一个跟随技术发展的教学环境，提高学生掌握知识、技能的效率，优化教学过程、提高教学质量，调动学生的学习积极性，突破教学的重点、难点的作用。与此同时，虚拟现实技术还结合了益智游戏、情景化学习、协作学习、远程教育等多种手段，

能够有效解决许多以前根本无法解决的教育问题。例如，通过 VR进行教育教学活动，在实现人与机器的交流、通过网络进行人与人之间交流的同时，还能够在某种程度上实现教学的游戏性，让教育真正做到寓教于乐，从而有效地激发学生的学习兴趣，进而更为高效地创建学习情景，支持合作，促进交流、知识表达及应用，进而构建优良的学习环境。因此，如果能将虚拟现实技术运用于现代化教育教学中，其效果将是难以言喻的。虚拟现实教学有如下的特点：

- 逼真的三维虚拟环境

能够使学生置身于虚拟的 3D学习环境中，体验一般实训中难以观察到的各部件以及重要部件独立工作的原理，让学生能够自己动手操作仪器，更好地了解仪器设备的构造、安装与操作。

- 可交互性强

让学生与学生、学生与老师之间可以像正常教学那样互动的同时，体验平时不可能体验到的情境，同时可以利用并行虚拟现实系统的共享性使所有学生可以共享资源，激发学习兴趣，提高学生的自主学习能力和操作能力。

- 不受时间和空间的限制

可以在任意的时间和地方利用 VR教学设备学习，学习内容也可以根据自己的兴趣和天分而定。与传统的教学方式相比，节约了时间，提高了学生的学习效率，同时充分挖掘了学生的兴趣爱好。

- 支持远程应用及管理

VR教育系统的平台采用服务器与用户端的模式。学生和老师通过用户端申请账号和密码进行登陆，然后访问服务器，进行相应数据的传输。学生和老师都可以在设备上学习，实现近乎真人“面对面”的教学，这样的教学方式比寻求家教更加方便，更加高效。

由于虚拟现实技术能够创建与现实社会类似的环境，进而解决学习媒体的情景化及自然交互性的要求，因此其在教育领域内有着极其广阔的应用前景。

在教育理论体系中，学习理论是处于核心地位的。虚拟现实技术在教育中的应用也是在一定的学习理论指导下进行的。虚拟现实技术本身是对客观对象的模拟，所构建的学习环境与实际生活情境相关。因此，如果把虚拟现实技术与教育有机地结合，教育与社会需求之间的差距将得以缩小，这将对未来的教育产生深远影响。

在教学实践中，亲身经历、亲身感受比空洞抽象的说教更具说服力，主动地去交互与被动的观看有着质的差别。因此，虚拟现实技术可以广泛用于学习情景的创设，增加学习内容的形象性和趣味性，进而实现模拟训练。通过 VR进行学习和教育，不光可以减少现实空间中某些训练操作的困难和危险，更可以使训练造价得到大幅度降低。因此，虚拟现实技术将是继多媒体、计算机网络之后，在教育领域内最具有应用前景的一项技术[68]。

4.6医学领域

□ “VR+”电子病历系统

电子病历是一种以电子化的形式记录患者信息的系统，包含着诊断记录、检查结果、手术记录、医嘱等等，便于存储患者的病情和治疗记录等信息。现行的电子病历系统仅仅是纸质版病历的电子化，其内容和功能方面并未有较大突破。随着信息化建设的加快，未来的病历系统应该是一个具有可视化、可交互功能的系统。利用 VR技术，将每位用户的器官三维模型上传到病历系统服务器，与用户的身份、病情等其他信息相匹配，并与相关的 VR设备连接，便可建立基于虚拟现实的电子病历系统。该系统可以包括用户的身体三维模型、诊断记录、手术记录、医嘱等，患者可以通过移动设备以扫描二维码的形式访问该系统，并利用VR设备浏览携带自身健康信息的三维模型。用户还可以在医生的指导下，利用数据手套或者其他的交互工具进行简单的人机交互，以直观明了的方式查看各种信息的图解、医嘱。由于该系统具有便捷、直观的特点，可以使患者全方位了解自己的健康状况，有利于患者的情绪稳定，还能促进医患之间的交流，尽可能减少医患纠纷。

□ “VR+”医学教学

传统的医学教学方式主要是通过板书、幻灯片、视频等二维形式呈现给学生，由于这些授课方式缺乏立体感和交互功能，学生们难以在短时间内快速领会到组织的空间毗邻关系和解剖关系。VR技术可以创造一个虚拟的三维课堂，将所需的课程内容经过计算机处理，加入必要的讲解，渲染之后发布到 VR设备中。课堂上的虚拟现实会使抽象的内容更加具体、专业知识更加有条理，全方位地展示教学内容，VR+教学的使用与课堂上讲义的使用相比，在学习性能和效率方面都有更好的效果。在手术模拟训练方面，虚拟现实技术也会发挥重要作用。手术操作技能练习的现行方式都是“师傅带徒弟”或者观看二维视频，针对的目标一般为尸体、动物和模拟人，这样就会导致成本高昂、针对性差、重用度低，而且患者必须承受医生成长的代价。如果对真实人体进行静态多源采集，并通过几何、物理、生理和建模，构建数字化人体，再将其作为虚拟手术平台，不仅可以替代那些成本较高的练习素材，也能大大提高练习效果。这种以“VR+医疗”为基础的虚拟手术平台让实习医生的手术训练减少对动物、尸体的依赖，并且尽可能地降低以实际患者作为手术实习对象所存在的风险。更加重要的是，虚拟现实的训练平台可以提供高度仿真的触觉反馈，使得训练素材甚至优于尸体和动物。可以预见，虚拟手术训练平台的应用将会给医学领域的人才储备发挥重要作用。

□ “AR+”外科手术导航

现行的外科手术导航都是通过 X光、B超、CT等手段，以二维显示器显示信息并提供指导，这类方法所呈现的患者信息缺失厉害，可读性差，特别是组织结构的空间位置和毗邻关系，缺乏立体感。医生往往依靠经验和直觉来判断病情，而且对于实施外科手术的医生往往会带来手眼不协调和深度感知丢失等问题。增强现实技术能够通过在原位叠加三维信息来克服这些难题，该技术在术中导航的巨大潜力。在进行影像分析时，利用三维重建能更加准确的判断病灶，加快诊断效率；在面对复杂的解剖关系时，利用 3D打印模型可以更方便的进行术前规划，并且能减少造影剂的使用剂量，将潜在的肾病风险降到最低。在手术过程中，增强现实的虚拟三维模型不会阻挡真实的环境，而是将虚拟信息叠加在真实环境

上，在不干扰原始任务的情况下提供直观的指导。由于具有直观性，增强现实可以在手术中简化操作程序，并做出精确判断，这将减少麻醉时间，还能降低手术不良后果的发生的几率。实时交互的功能也允许外科医生充满血渍的手来进行手势操控，便捷地转换自己需要的视野或者其他的指导。因此，在像外科手术这样的敏感操作，选择 AR技术能够提供更安全的操作环境和更高的手术效率。研究表明，AR系统在安全可靠等方面已足以与传统导航技术相媲美，其精确度和安全性也能满足常规临床实践。

□ “AR+”微创手术

微创手术是现代医学进步的产物，具有创口小、痛苦小、恢复快等优点。但是微创手术应用至今，最为关键的成像问题却一直没有得到改变。就心血管疾病而言，实施这种介入治疗的最大难题之一是血管外科医生必须通过二维的血管造影图像联想出三维血管形状。因此改善血管成像是微创手术成功的一个关键要素，血管的可视化与手术工具的空间位置和方向对微创手术的成功都有着重要影响。现行的微创手术成像主要依靠内窥镜成像、B 超成像和 X 光成像。内窥镜是通过表面成像，皮下信息缺失，由于体内环境所限导致视野狭窄；B超是通过断面回波成像，该成像方式往往有很强的噪声干扰，得到的仅仅是模糊的二维图像，缺乏三维信息；X光技术是借助于透视投影成像得到骨骼或组织的投影图，这种图像往往存在多种组织成像交叠，同样缺少三维信息。总体看来，问题还是在于成像质量差、精准定位难、空间信息少。增强现实技术的使用有望解决这些问题。在介入手术中，增强现实技术通过三维重建解剖结构、肿瘤靶点，使介入过程更加简单、自信地进行，具有更高的靶向精度。可以预见，增强现实将会利用可视化信息来代替传统的二维图像，使得微创手术应用更加广泛[69]。

AMiner

5趋势篇

领域技术分析系统（http://trend.aminer.cn）可以对所选期刊和会议中的科技情报进行深入挖掘，对技术趋势、国际趋势及机构趋势等方面进行分析。

□ 技术趋势

技术趋势分析如图 5-1所示。技术趋势分析描述了技术的出现、变迁和消亡的全过程，可以帮助研究人员理解领域的研究历史和现状，快速识别研究的前沿热点问题。图中每条色带表示一个话题，其宽度表示该术语在当年的热度，与当年该话题的论文数量呈正相关，每一年份中按照其热度由高到低进行排序。通过技术趋势分析可以发现当前该领域的热点研究话题 TOP10是：Virtual Reality、Augmented Reality、Social Media、Social Interaction、Interaction Design、Mobile Device、Social Network、Ubiquitous Computing、Mobile Phone、Interaction Technique。

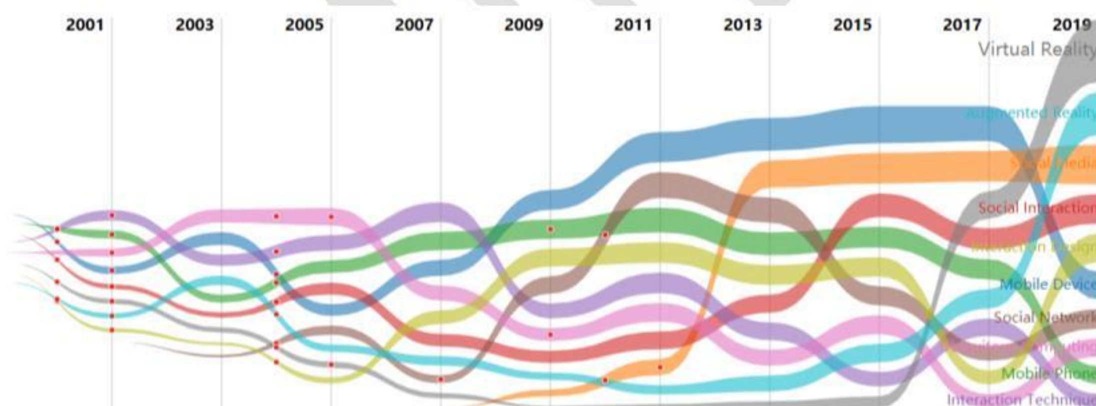


图 5-1 人机交互技术趋势

□ 国家趋势

国家趋势分析如图 5-2所示。图中每条色带表示一个国家，其宽度表示该国家在当年的研究热度，与当年该国论文数量呈正相关，每一年份中按照其热度由高到低进行排序。国家趋势分析显示当前人机交互领域研究热度 TOP10 的国家分别是：United States、United Kingdom、Germany、Canada、China、Japan、South Korea、Australia、France、Netherlands。

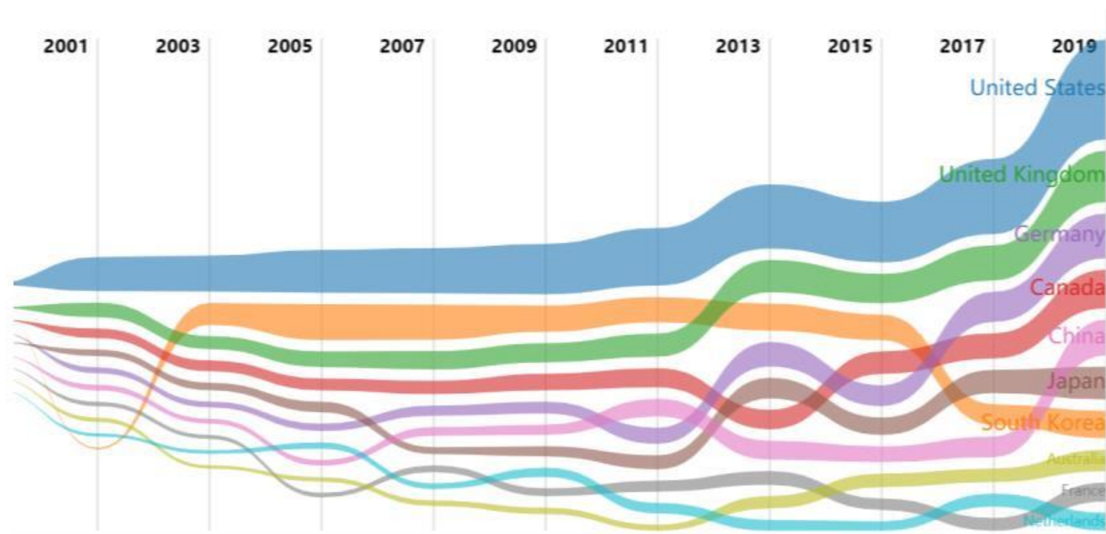


图 5-2 人机交互国家趋势

□ 机构趋势

机构趋势分析如图 5-3所示。图中每条色带表示一个机构，其宽度表示该机构在当年的研究热度，与当年该机构论文数量呈正相关，每一年份中按照其热度由高到低进行排序。机构趋势分析显示当前人机交互领域研究热度 TOP10 的机构分别是： Carnegie Mellon University、 Washington College、University of California、Stanford University、University of Michigan、Massachusetts Institute of Technology、Georgia Institute of Technology、Cornell University、Seoul National University、Yonsei University。

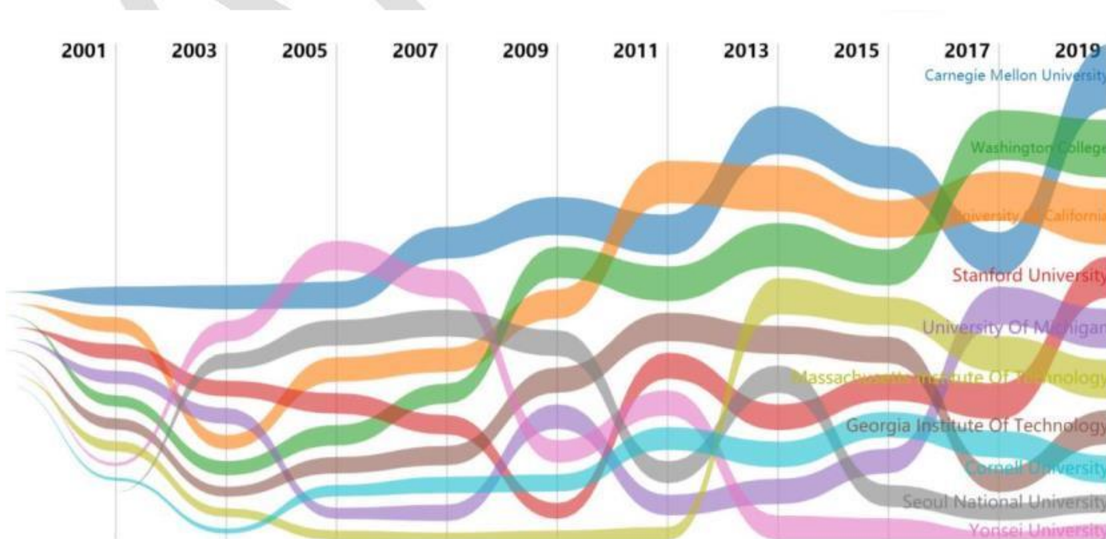


图 5-3 人机交互机构趋势

AMiner

6结语

最近十年，是人机交互向着自然交互蓬勃发展的十年。毋庸置疑，计算机是人类最伟大的发明之一，其作用从科学计算工具迅速发展为信息处理和信息交互工具，人机交互技术的变革对计算机技术的发展起到了引领作用。以鼠标为标志的图形用户界面的产生，一改规范命令与计算机交互的命令行界面模式，图形用户界面提供了普通人与计算机便捷交互的工具和方法，让计算机从实验室走进办公室、走入家庭。随着触屏技术的发展，图形用户界面中的鼠标被手指的天然指点所取代，手机成为更多人方便使用的随身掌上工具。更少地依赖操控工具，发展和使用成本更小的自然交互技术，一直是人机交互研究的追求，最近十年，随着感知和计算技术的进步，自然交互技术创新层出不穷，并能迅速应用于新型产品中。《麻省理工科技评论》总结和评论了人机交互领域的突破技术，这些突破技术分为三大类：支持自然动作的感知技术、面向穿戴的新型终端以及基于语音识别的对话交互，这些技术突破对人机交互的发展影响深远。

人体动作蕴含丰富的语义，动作交互技术一方面需要感知技术的进步，另一方面需要发现或设计有明确交互语义的动作。如今，二维表面上，多指触摸动作在触屏上已普遍可用；三维空间中，嵌入了深度摄像头的手持和固定设备，能比较准确识别人的姿态和动作，做出响应。不同于人脸识别等目标明确的视觉识别任务，动作交互不仅要求视觉识别的准确度，更需要研究基于交互任务的动作表达的自然性与一致性，除了动作语义很直白的动作游戏，三维动作交互尚缺少普遍认知和接受的交互动作语义，而无论二维还是三维，手势的不可见性是动作交互的主要难题。

穿戴（wearable）取代手持（handheld）曾是前几年的一个革命口号，目前，市场上的确出现了一定规模的新产品，但穿戴仍是处于补充的地位。穿戴设备中，手环设备基本只有健康和活动检测功能，智能手表可以算做创新终端，但作为缩小版的手机，由于交互界面的缩小和操作方式的限制（通常是小界面上双手参与操作），其承载功能也较手机缩减很多。虚拟现实 /增强现实的一个理想载体是头戴式设备，最近几年，多款智能眼镜产品面世，较之前笨重的头盔轻便了许多，逼真的虚拟场景和准确的现实对象识别信息都可以清晰呈现在眼前，并在特定领

域开拓着增强体验的应用；然而，智能眼镜尚缺少与其三维真实显示匹配的准确的自然输入技术，以及从眼手绑定在手机上转变到眼手分离的眼镜设备上时，尚未建立起相应的交互模式。

自然语言对话式交互得益于大数据和智能技术的进步，多语言的自然语音识别技术在用户终端上都达到了很高的可用水平，同时，语音识别超越文本输入方式，成为智能软件助理的重要技术，近两年，更是有基于语音接口的家居产品如雨后春笋般出现。然而，语音用户界面的局限也是显而易见的，相对并行模式的视觉通道，串行模式的语音通道的带宽显然窄得多，出声的使用方式在很多场合是不合适的。但作为一种可用的自然交互技术，有效提升了用户体验。

人机交互作为终端产品引领技术，其作用已经为产业界所普遍认识，多种自然交互技术和新型交互终端相继面世，但图形用户界面仍是交互的主导模式。计算无所不在，人机交互的研究和开发空间很大，自然高效的交互是发展趋势，需要综合地探索自然交互技术的科学原理，建立明确的优化目标，结合智能技术，发展高效可用的自然交互技术。

AMiner

参考文献

[1] Card, Stuart K. The psychology of human-computer interaction[M]. *Crc Press*, 2018.

[2] 袁保宗, 阮秋琦, 王延江, 刘汝杰, 唐晓芳. 新一代(第四代)人机交互的概念框架特征及关键技术[J]. 电子学报, 2003(S1): 1945-1954.

[3] 人机交互发展历史. https://wenku.baidu.com/view/faeda09d0a1c59eef8c75fbfc77da26924c596ee.html.

[4] 易鑫. 自然文本输入中的贝叶斯推理方法[D]. 清华大学, 2018.

[5] Johnson E A. Touch display--a novel input/output device for computers[J]. Electronics Letters, 1965, 1(8): 219-220.

[6] 朱莉. 高可靠电阻式触摸屏的研究与实现[D]. 东南大学, 2015.

[7] Pieraccini R. The voice in the machine: building computers that understand speech[M]. *MIT Press*, 2012.

[8] 周盼. 基于深层神经网络的语音识别声学建模研究 [D]. 中国科学技术大学 , 2014.

[9] 张建华. 基于深度学习的语音识别应用研究[D]. 北京邮电大学, 2015.

[10]刘浩杰, 杜利民. 语音合成技术的发展与展望 [J]. 微计算机应用 , 2007(07): 726-730.

[11]张斌, 全昌勤, 任福继. 语音合成方法和发展综述 [J]. 小型微型计算机系统 , 2016, 37(01): 186-192.

[12]阎裕康. 自然动作输入的“编码-解码”优化方法[D]. 清华大学, 2020.

[13]郭小爽 . 人机交互中的动态手势识别及应用研究 [D]. 西安电子科技大学 , 2014.

[14]武霞 , 张崎 , 许艳旭 . 手势识别研究发展现状综述 [J]. 电子科技 , 2013, 26(06):171-174.

[15]方美玲. 随机森林算法在身体姿势识别中的应用研究 [D]. 西安电子科技大学, 2016.

[16]刘昕. 基于眼动的智能人机交互技术与应用研究[D]. 南京大学, 2019.
[17]Jens, et al., The Office of the Future: Virtual, Portable, and Global. 2019.
[18]Rhee, T., et al., Augmented Virtual Teleportation for High-Fidelity Telecollaboration. IEEE Transactions on Visualization and Computer Graphics, 2020. 26(5): p. 1923-1933.
[19]Ma, C., et al. An adaptive sketching user interface for education system in virtual reality. in 2009 IEEE International Symposium on IT in Medicine & Education. 2009.
[20]Walker, J.K. Decoder-Assisted Typing using an HMD and a Physical Keyboard. 2016.
[21]Walker, J., et al. Efficient Typing on a Visually Occluded Physical Keyboard. in Chi Conference. 2017.
[22]Kim, S. and G.J. Kim, Using keyboards with head mounted displays, in Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry. 2004, Association for Computing Machinery: Singapore. p. 336–343.
[23]Bovet, S., et al. Using Traditional Keyboards in VR: SteamVR Developer Kit and Pilot Game User Study. in 2018 IEEE Games, Entertainment, Media Conference (GEM). 2018.
[24]Grubert, J., et al. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018.
[25]McGill, M., et al., A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays, in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015, Association for Computing Machinery: Seoul, Republic of Korea. p. 2143–2152.

AMiner

[26]Jiang, H., et al. HiKeyb: High-Efficiency Mixed Reality System for Text Entry. in 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). 2018.

[27]Bowman, D., C. Rhoton, and M. Pinho, Text Input Techniques for Immersive Virtual Environments: An Empirical Comparison. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2002. 46.

[28]Gonzalez, G., et al., Evaluation of Text Input Techniques in Immersive Virtual Environments. 2009. p. 109-118.

[29]Grubert, J., et al. Effects of Hand Representations for Typing in Virtual Reality. in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018.

[30]Knierim, P., et al., Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018, Association for Computing Machinery: Montreal QC, Canada. p. Paper 345.

[31]Lin, J.-W., et al., Visualizing the keyboard in virtual reality for enhancing immersive experience, in ACM SIGGRAPH 2017 Posters. 2017, Association for Computing Machinery: Los Angeles, California. p. Article 35.

[32]Speicher, M., et al., Selection-based Text Entry in Virtual Reality, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018, Association for Computing Machinery: Montreal QC, Canada. p. Paper 647.

[33]Min, K., Text Input Tool for Immersive VR Based on 3 × 3 Screen Cells. 2011. 778-786.

[34]Chen, S., et al., Exploring Word-gesture Text Entry Techniques in Virtual Reality. 2019. 1-6.

[35]Xu, W., et al. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. in 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2019.

[36]Yu, C., et al., Tap, Dwell or Gesture? Exploring Head-Based Text Entry Techniques for HMDs, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017, Association for Computing Machinery: Denver, Colorado, USA. p. 4479–4488.

[37]Majaranta, P., U.-K. Ahola, and O. Špakov, Fast gaze typing with an adjustable dwell time, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009, Association for Computing Machinery: Boston, MA, USA. p. 357–360.

[38]Bowman, D.A.L., Vinh Q.;Campbell, Joshua M., Pinch Keyboard: Natural Text Input for Immersive Virtual Environments.

[39]Mehring, C., et al. KITTY: keyboard independent touch typing in VR. in IEEE Virtual Reality 2004. 2004.

[40]Whitmire, E., et al., DigiTouch: Reconfigurable Thumb-to-Finger Input and Text Entry on Head-mounted Displays. 2017. 1(3 %J Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.): p. Article 113.

[41]Wong, P.C., K. Zhu, and H. Fu, FingerT9: Leveraging Thumb-to-finger Interaction for Same-side-hand Text Entry on Smartwatches, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018, Association for Computing Machinery: Montreal QC, Canada. p. Paper 178.

[42]Xu, Z., et al., TipText: Eyes-Free Text Entry on a Fingertip Keyboard, in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. 2019, Association for Computing Machinery: New Orleans, LA, USA. p. 883–899.

[43]Xu, Z., et al., BiTipText: Bimanual Eyes-Free Text Entry on a Fingertip Keyboard, in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020, Association for Computing Machinery: Honolulu, HI, USA. p. 1–13.

[44]Gupta, A., et al., RotoSwype: Word-Gesture Typing using a Ring, in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019, Association for Computing Machinery: Glasgow, Scotland Uk. p. Paper 14.

[45]Gugenheimer, J., et al., FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality, in Proceedings of the 29th Annual Symposium on User Interface Software and Technology. 2016, Association for Computing Machinery: Tokyo, Japan. p. 49–60.

[46]Kim, Y.R. and G.J. Kim. HoVR-Type: Smartphone as a typing interface in VR using hovering. in 2017 IEEE International Conference on Consumer Electronics (ICCE). 2017.

[47]Lu, Y., et al., BlindType: Eyes-Free Text Entry on Handheld Touchpad by Leveraging Thumb's Muscle Memory. 2017. 1(2 %J Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.): p. Article 18.

[48]Rajanna, V. and J.P. Hansen, Gaze typing in virtual reality: impact of keyboard design, selection method, and motion, in Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. 2018, Association for Computing Machinery: Warsaw, Poland. p. Article 15.

[49]Ma, X., et al., Combining Brain-Computer Interface and Eye Tracking for High-Speed Text Entry in Virtual Reality, in 23rd International Conference on Intelligent User Interfaces. 2018, Association for Computing Machinery: Tokyo, Japan. p. 263–267.

[50]Hsieh, Y.-T., et al., Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses, in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016, Association for Computing Machinery: San Jose, California, USA. p. 4203–4215.

[51]Dudley, J.J., K. Vertanen, and P.O. Kristensson, Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. 2018. 25(6 %J ACM Trans. Comput.-Hum. Interact.): p. Article 30.

AMiner

[52]Kim, J., W. Delamare, and P. Irani, ThumbText: Text Entry for Wearable Devices Using a Miniature Ring, in Proceedings of the 44th Graphics Interface Conference. 2018, Canadian Human-Computer Communications Society: Toronto, Canada. p. 18–25.

[53]Jiang, H.W., Dongdong; Zhang, Zhenliang; Chen, Feng., HiFinger: One-Handed Text Entry Technique for Virtual Environments Based on Touches between Fingers. Sensors 19, no. 14: 3063., 2019.

[54]BlueTap — The Ultimate Virtual-Reality (VR) Keyboard.

[55]Prätorius, M., et al., DigiTap: an eyes-free VR/AR symbolic input device, in Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology. 2014, Association for Computing Machinery: Edinburgh, Scotland. p. 9–18.

[56]Ogitani, T., et al. Space Saving Text Input Method for Head Mounted Display with Virtual 12-key Keyboard. in 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA). 2018.

[57]Yu, D., et al., PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. IEEE Transactions on Visualization and Computer Graphics, 2018. 24(11): p. 2927-2935.

[58]Xu, W., et al., RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. IEEE Transactions on Visualization and Computer Graphics, 2019. 25(5): p. 1991-2001.

[59]Yanagihara, N. and B. Shizuki, Cubic Keyboard for Virtual Reality, in Proceedings of the Symposium on Spatial User Interaction. 2018, Association for Computing Machinery: Berlin, Germany. p. 170.

[60]阿钦蒂亚 • K.鲍米克. 实感交互:人工智能下的人机交互技术. 机械工业出版社. 2018 年 7月第 1版.

[61]史元春. 信息无障碍中的智能交互技术 [J]. 中国计算学会通讯, 2020, 16(3): 8-10.

[62]在人机交互方面手机厂商有什么改变和创新. https://baijiahao.baidu.com/s?id=1657774827501742777&wfr=spider&for=pc.
[63]赵默涵. 车载娱乐系统人机交互设计研究[D]. 吉林大学, 2018..
[64]朱文凯, 王佳. 浅谈智能化的单兵系统[J]. 国防技术基础, 2008(12): 52-55.
[65]傅耀威, 孟宪佳, 王涌天. 可穿戴移动终端的多感官人机交互技术发展现状与趋势[J]. 科技中国, 2017(07): 12-15.
[66]智能家居的人机交互方式有哪些. https://zhuanlan.zhihu.com/p/26845446.
[67]基于体感技术的人机交互设计在游戏领域的应用[J]. 艺术科技, 2017, 30(10): 103.
[68]人机交互在教育领域的应用. https://blog.csdn.net/qq_40531073/article/details/89500087.
[69]石晓卫, 苑慧, 吕茗萱, 蔡佳惠, 张先增. 虚拟现实技术在医学领域的研究现状与进展[J/OL]. 激光与光电子学进展: 1-17.

AMiner

版权声明

AMiner 研究报告版权为 AMiner 团队独家所有，拥有唯一著作权。AMiner 咨询产品是 AMiner团队的研究与统计成果，其性质是供用户内部参考的资料。

AMiner研究报告提供给订阅用户使用，仅限于用户内部使用。未获得 AMiner 团队授权，任何人和单位不得以任何方式在任何媒体上（包括互联网）公开发布、复制，且不得以任何方式将研究报告的内容提供给其他单位或个人使用。如引用、刊发，需注明出处为“AMiner.org”，且不得对本报告进行有悖原意的删节与修改。

AMiner研究报告是基于 AMiner团队及其研究员认可的研究资料，所有资料源自AMiner后台程序对大数据的自动分析得到，本研究报告仅作为参考，AMiner 团队不保证所分析得到的准确性和完整性，也不承担任何投资者因使用本产品与服务而产生的任何责任。

AMiner

考研资料：数学、英语、政治、管综、西综、法硕等（整合各大机构）
英语类：四六级万词班专四专八雅思等
财经类：初级会计、中级会计、注册会计、高级会计、税务师、会计实操、证券从业、基金从业、资产评估、初级审
公务员：国考、省考、事业单位、军队文职、三支一扶微信 2270291360
银行：银行招聘、笔试、面试
教师资格：小学、中学、教师招聘面试
建筑：一建、二建、消防、造价
法考：主观题、客观题
多平台网课：涵盖职场、办公技能、编程、文案写作、情感心理、穿搭技巧、理财投资健身减肥摄影技术等优质内容
精选资料：Excel 教程、PPT 模板、简历模板、PS 教程、PPT 教程、素描、烹饪、小语种、CAD 教程、PR 教程、UI 课程、自媒体、写作、计算机二级、钢琴、Python、书法、吉他、kindle 电子书、演讲.....持续更新中...
押题：提供考前冲刺押题（初级会计、中级会计、注册会计、一建、二建、教资、四六级、证券、基金、期货等等），麻麻再也不用担心我考不过了。
资料领取微信：1131084518
行业报告：20000 份+持续更新

英语四六级备考资料	计算机二级备考资料	150 所高校考研专业课资料
两小时搞定毛概马原思修近代史纲	高数(微积分)+线性代数+概率论	素描 0 基础入门教程
教师资格证全套备考资料	普通话考试资料礼包	书法教程微信 2270291360
大学生英语竞赛备考资料	大学生数学竞赛备考资料	1000 份各行业营销策划方案合集
挑战杯/创青春/互联网+竞赛资料	电子设计竞赛必备资料	街舞 0 基础入门教程
托福雅思备考资料	大学物理学科攻略合集	动漫自学教程
SCI 最全写作攻略	TEM4/TEM8 专四专八备考资料	教师资格证面试试讲万能模板
360 份精美简历模板	数学建模 0 基础从入门到精通	100 套快闪 PPT 模板
Vlog 制作最全攻略	超强 PR 模板	42 套卡通风 PPT 模板
PS 零基础教程微信 1131084518	PS 高级技能教程	63 套酷炫科技 PPT 模板
好用到极致的 PPT 素材	128 套中国风 PPT 模板	32 套 MBE 风格 PPT 模板
327 套水彩风 PPT 模板	295 套手绘风 PPT 模板	54 套毕业答辩专属 PPT
196 套日系和风 PPT 模板	82 套文艺清新 PPT 模板	57 套思维导图 PPT 模板
163 套学术答辩 PPT 模板	53 套北欧风 PPT 模板	34 套温暖治愈系 PPT 模板
118 套国潮风 PPT 模板	30 套仙系古风 PPT	126 套黑板风 PPT 模板
114 套星空风格 PPT 模板	192 套欧美商务风 PPT 模板	42 套绚丽晕染风 PPT
50 套精美 INS 风 PPT 模板	56 套水墨风 PPT 模板	137 套清爽夏日风 PPT 模板
98 套森系 PPT 模板	25 套简约通用 PPT 模板	记忆力训练教程
300 套教学说课 PPT 模板	123 套医学护理 PPT 模板	AE 动态模板微信 2270291360
毕业论文资料礼包	教师资格证重点笔记+易错题集	表情包制作教程
吉他自学教程（送 6000 谱）	钢琴自学教程（送 1000 谱）	区块链从入门到精通资料
2000 部 TED 演讲视频合集	Excel 从入门到精通自学教程	单片机教程
230 套可视化 Excel 模板	1000 款 PR 预设+音效	1000 份实习报告模板
手绘自学教程微信 1131084518	单反从入门到精通教程	人力资源管理师备考资料
英语口语自学攻略	粤语 0 基础从入门到精通教程	证券从业资格证备考资料
日语自学教程	韩语自学教程	PHP 从入门到精通教程
法语学习资料	西班牙语学习资料	炒股+投资理财从入门到精通教程
全国翻译专业资格考试备考资料	BEC 初级+中级+高级全套备考资料	大数据学习资料
SPSS 自学必备教程	Origin 自学必备教程	会计实操资料
LaTeX 全套教程+模板	EndNote 教程+模板	小提琴 0 基础入门自学教程
GRE 超全备考资料	200 份医学习题合集	司考备考资料

思维导图教程＋素材	20000 份行业研究报告	一级造价工程师备考资料

上万 GB 教学资料（均全套,非杂乱）免费领取微信 1131084518

《闪电式百万富翁》实战版+升级版

易经+道德经+易学名师全集+风水学+算命学+起名+++等等（全套 1000 多 GB）

心理学+NLP 教练技术+精神分析+亲子家庭教育+催眠+++等等（更新超 2000GB）

大学-已更新至 9333 个课程+高中+初中+小学-全套资料（超过 2 万 GB）

陈安之	曾仕强	马云	杜云生	翟鸿燊	刘一秒	俞凌雄
王健林	余世维	雷军	周文强	安东尼罗宾	董明珠	李嘉诚
徐鹤宁	冯晓强	李践	刘克亚	罗伯特清崎	戴志强	李伟贤
苏引华	史玉柱	李强	俞敏洪	杰亚伯拉罕	周鸿祎	唐骏
梁凯恩	陈永亮	傅佩荣	贾长松	易发久	李彦宏	湖畔大学
李开复	聂枭	悟空道场	魏星	姬剑晶	其他名师全集	其他资料下载
王兴	王智华	智多星	陈文强	周导		微信 2270291360
泡妞	撩汉	泡仔	房中术	性福课	驻容术	泡妞撩汉性福合集

注：太多了，无法全部一一列出。。。

全套专题系列【微信 1131084518】

记忆力训练	形象礼仪	健康养生	企业管理	沟通技巧
演讲与口才	经理修炼	MBA 讲座	时间管理	战略经营
企业文化	销售心理	管理素质	国学讲座	执行力
团队管理	领导艺术	员工激励	潜能激发	谈判技巧
绩效管理	薪酬管理	43 份直销制度	电话销售	人力管理
客户服务	创业指南	市场营销	餐饮管理	保险讲座
品牌营销	酒店管理	汽车 4S 店	众筹资料	销售技巧

兴趣爱好：	钓鱼教程	魔术教学	炒股教学	美术教学	书法教学	
音乐乐器：	萨克斯教学	电子琴教学	小提琴	古筝教学	钢琴教学	
	吉他教学					
体育运动：	篮球教学	足球教学	羽毛球教学	乒乓球教学	太极拳教学	
	围棋教学	高尔夫球				
生活实用：	插花教学	茶艺-茶道	唱歌教学	单反相机摄影	毛线编织	
	小吃+美食					暗号：666
语言学习：	英语					免费领取资料微信
电脑 IT：	办公 office	PS 美工教学				1131084518

微信1131084518
撩汉liaohan.net
最好资源zuihaoziyuan.com



如果群里报告过期

请加微信联系我索取最新

1、每日微信群内分享7+最新重磅报告;

2、每日分享当日华尔街日报、金融时报

3、如果看到群里报告过期了，请扫码联系

4、行研报告均为公开版，权利归原作者所有，仅分发做内部学习

扫一扫二维码后台回复 加群

加入“研究报告”微信报告群。

